(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/046220 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IL02/00904

(22) International Filing Date: 11 November 2002 (11.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/993,398 26 November 2001 (26.11.2001) US
10/201,605 24 July 2002 (24.07.2002) US

(71) Applicant *(for all designated States except US)*: **COMPUGEN LTD.** [IL/IL]; 72 Pinchas Rosen Street, 69 512 Tel Aviv (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LEVANON, Erez** [IL/IL]; 73 Menachem Begin Street, 49 732 Petach Tikva (IL). **POLLOCK, Sarah** [IL/IL]; 14b Almagor Street, 69 187 Tel Aviv (IL). **NEMZER, Sergey** [IL/IL]; 5 HaRenanim Street, 52 595 Ramat Gan (IL). **SHOSHAN, Avi** [IL/IL]; 2/4 Rechavat Har Sinai, 82 000 Kiryat Gat (IL). **KHOSRAVI, Rami** [IL/IL]; 36/1 HaLevi Street, 46 490 Herzlia (IL). **WALACH, Shira** [IL/IL]; 40 Bnei Brith Street, 45 265 Hod HaSharon (IL). **LEVINE, Zurit** [IL/IL]; 5 HaKozrim Street, Herzlia 46 360 (IL). **BERNSTEIN, Jeanne** [IL/IL]; 23 HaRimon Street, 40 300 Kfar Yona (IL). **DAHARI, Dvir** [IL/IL]; 17 HaZanchanim Street, 69 270 Tel Aviv (IL). **WASSERMAN, Alon** [IL/IL]; 5 HaAmoraim Street, Tel Aviv 69 207 (IL). **ROTMAN, Galit** [IL/IL]; 5 Yair Shtern Street, 46 412 Herzlia (IL).

(74) Agent: **G. E. EHRLICH (1995) LTD.**; 28 Bezalel Street, 52 521 Ramat Gan (IL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/046220 A1

(54) Title: METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE TRANSCRIPTS AND METHODS, KITS AND ARRAYS UTILIZING SAME

(57) Abstract: A method of identifying putative naturally occurring antisense transcripts is provided. The method is effected by (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database, thereby identifying putative naturally occurring antisense transcripts.

230

METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE TRANSCRIPTS AND METHODS, KITS AND ARRAYS UTILIZING SAME

BACKGROUND AND FIELD OF THE INVENTION

The present invention relates to the field of naturally occurring, antisense transcripts. More particularly, the present invention relates to methods of identifying naturally occurring antisense transcripts, databases storing polynucleotide sequences encoding identified naturally occurring antisense transcripts, oligonucleotides derived therefrom and methods and kits utilizing same.

Naturally occurring antisense RNA transcripts are endogenous transcripts, which exhibit complementarity to sense transcripts of which are typically of a known function. It has been established that these endogenous antisense transcripts play an important role in regulating prokaryotic gene expression and are increasingly implicated as involved in eukaryotic gene regulation.

Cis-encoded antisense transcripts are encoded by the same locus as the sense transcripts and are transcribed from strand of DNA opposite to that encoding the sense transcript; as such, cis encoded antisense transcripts are typically completely complementary with a portion of the sense transcript.

Trans-encoded antisense transcripts are by contrast, transcripts, which are encoded on a different locus and as such, may display only partial complementarity with a sense transcript.

Natural antisense RNAs were first described in prokaryote studies, which suggested that such transcripts play a role in gene expression regulation. Prokaryotic antisense transcripts are widely distributed and are involved in the control of numerous biological functions including transposition, plasmid replication, incompatibility and conjugation. In prokaryotes, antisense transcripts are typically involved in down-regulation of sense transcript

expression, although involvement in positive regulation was also suggested [reviewed in Wagner EG. and Simons RW. (1994) Annu. Rev. Microbiol. 48:713-742].

The first example of transcription from both strands of eukaryotic DNA was illustrated in human and mouse mitochondrial genes [Anderson S. et al. (1981) Nature 290:457-465 and Bibb MJ. et al. (1981) Cell 26:167-180]. Since then, examples of antisense transcripts have been documented in a variety of organisms including viruses, slime molds, insects, amphibians and birds as well as mammals. It is thought that these antisense RNAs are involved in extremely diverse biological functions, such as, hormonal response, control of proliferation, development, structure, viral replication and others. Some antisense RNAs are conserved between species suggesting that these antisense RNAs are not fortuitous but rather play an important role in gene expression regulation [Kidny MS. et al. (1987) Mol. Cell Biol. 7:2857-2862, Nepveu A. and Marcu KB. (1986) EMBO J. 5:2859-2865 and Bentley DL. et al. (1986) Nature 321:702-706].

Antisense transcripts can also encode proteins. Examples for protein encoding antisense transcripts include *rev-ErbAx* [Lazar MA. (1989) Mol. Cell. Biol. 9:1128-1136], *gfg* [Kimelman D. et al. (1989) Cell 59:687-696] and *n-cym* [Armstrong BC. et al. (1992) Cell Growth Differ. 3:385-390]. Such antisense transcripts typically include a distinct open reading frame (ORF) and polyadenylation signal for cytoplasm transportation.

However, it is believed that most antisense transcripts play a role in gene expression regulation. This assumption is mostly based on spatial and/or temporal distributions of sense and antisense transcripts. Indeed, tissue distribution studies suggest that high levels of sense and antisense transcripts rarely occur together, as was exemplified for the *dopa decarboxylase* transcripts in *Drosophila* [Spencer CA. et al. (1986) Nature 322:279-281]. Additional studies demonstrated that changes in sense gene expression correlate with presence of antisense RNA. Furthermore, an inverse relationship between

levels of accumulation of sense and antisense transcripts such as has been reported for *α1 (I) collagen* transcripts in chondrocytes under chemotherapy has also been reported [Farrell CM. And Lukens LN. (1995) J. Biol. Chem. 270:3400-3408]. However, it will be appreciated that mutual expression of sense and their corresponding antisense transcripts is also reported and may involve a different mechanism of regulation.

Evidence for involvement of antisense-mediated gene regulation in the development of pathologies has also been presented. For example, endogenous antisense transcripts may be involved in regulation of the expression levels of the tumor suppressor gene WT1 observed in Wilm's tumors [Eccles MR. et al. (1994) Oncogene 9:2059-2063].

Natural antisense regulation of gene expression can be effected via one of several mechanisms.

Nuclear regulation

Nuclear regulation can be effected via several gene-processing pathways [reviewed in Vanhee-Brosollet C. and Vaquero C. (1998) Gene 211:1-9]

dsRNA-mediated DNA methylation - complementation between endogenous sense transcripts and antisense transcripts of sequences as short as 30 bp may initiate DNA-methylation, a well-established phenomenon in a number of organisms [Sharp A. (2001) Genes Dev. 15:485-490]. Methylation can be directed to different portions of an encoding region of the gene or to the promoter region. DNA methylation results in complete suppression of transcription probably by recruitment of histone deacetylases.

Transcriptional regulation – in which case antisense transcription hampers sense transcription. Such interference may involve the collision of two transcription complexes. Alternatively, interference may result from competition on an essential rate limiting transcription factor resulting in premature termination or in reduced elongation of transcription, the transcripts with the highest rate of transcription being predominant.

Post-transcriptional nuclear regulation – involves antisense intervention of either maturation and/or transport of the sense transcript to the cytoplasm. Alternatively, antisense transcripts displaying similar structural features to sense transcripts can bind proteins expected to interact with their sense counterparts, thereby depriving sense messengers from proteins necessary for their function.

Cytoplasmic regulation

Messenger stability –double stranded RNA may affect messenger stability via "RNA interference", which involves short segments of double stranded RNA (dsRNA) homologous in sequence to the silenced gene. These undersized segments, which are generated by a ribonuclease III cleavage of longer dsRNAs, can guide a single stranded target mRNA, via base pairing, to a multisubunit complex which participates in the degradation of the target mRNA. Alternatively, messenger stability may be affected by RNA degradation, which is mediated by double stranded RNA-directed Rnases.

Translation – masking the 3' untranslated region (UTR) and the polyA tail of the sense transcript is believed to modulate translation efficiency probably via direct or indirect interaction between 3'-proximal elements and upstream sequences or structures [reviewed in Jackson RJ. And Standart N. (1990) Cell 62:15-24].

Realizing the fundamental role antisense transcripts play in regulating sense transcription, stability and function, resulted in a number of attempts to systematically identify natural antisense transcripts. Accordingly, differential approaches were taken for exploring non-coding antisense RNA transcripts and antisense transcripts including an ORF. Although the latter carries ORF consensus parameters, uncovering antisense data from general sequence databases has proven to be a complicated task, as many of these sequences include an evolutionary conserved secondary structure rather than a conserved primary sequence, therefore primary sequence alignment methods are often not

very effective. Indeed, only a few attempts have been tried to date with only limited success.

Maziel's group [Chen JH. et al. (1990) Comput. Applic. Biosci. 6:7-18 and Le SY. et al (1990) Human Genome Initiative and DNA Recombination Vol. 1:127-136] has experimented with methods that look for regions of a genome with predicted RNA structures that are significantly more stable thermodynamically than random sequence of the same base composition. Although this approach detected a few highly structured non-coding RNAs, as well as few *cis*-regulatory structures, it appears that it is of limited use for large-scale applications.

Another approach examined coding dense genomes, having suspicious-looking large regions with little or no coding potential termed "gray holes" [Olivas WM. et al. (1997) Nucleic acids Res. 25:4619-4625]. Fifty nine gray holes were tested in the yeast genome. Northern analysis detected distinct transcripts from 15 of the gray holes. Only one transcript appeared to be a non-coding antisense transcript illustrating the low efficiency of this method.

There is thus a widely recognized need for, and it would be highly advantageous to have, methods of systematically identifying novel naturally occurring antisense molecules and methods of artificially generating and using same for detecting, quantifying and/or regulating sense transcripts, such as for example, mRNA transcripts associated with a pathological state.

SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts, the method comprising: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a

duplex with at least one sense-oriented polynucleotide sequence of the first database, thereby identifying putative naturally occurring antisense transcripts.

According to another aspect of the present invention there is provided a kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, the sequence region not being complementary with a naturally occurring antisense transcript.

According to yet another aspect of the present invention there is provided a kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA transcript of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest.

According to still another aspect of the present invention there is provided a method of designing artificial antisense transcripts, the method comprising: (a) providing a database of naturally occurring antisense transcripts; (b) extracting from the database criteria governing structure and/or function of the naturally occurring antisense transcripts; and (c) designing the artificial antisense transcripts according to the criteria.

According to further features in preferred embodiments of the invention described below the criteria governing structure and/or function of the naturally occurring antisense transcripts are selected from the group consisting of antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications,

protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

According to an additional aspect of the present invention there is provided a computer readable storage medium comprising a database including a plurality of sequences, wherein each sequence is of a naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the database further includes information pertaining to each sequence of the naturally occurring antisense transcripts, the information is selected from the group consisting of related sense gene, antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

According to still further features in the described preferred embodiments the database further includes information pertaining to generation of the database and potential uses of the database.

According to yet an additional aspect of the present invention there is provided a method of generating a database of naturally occurring antisense transcripts, the method comprising: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database so as to identify putative naturally occurring antisense transcripts; and (c) storing sequence information of the identified naturally occurring antisense

transcripts, thereby generating the database of the naturally occurring antisense transcripts.

According to still an additional aspect of the present invention there is provided a system for generating a database of a plurality of putative naturally occurring antisense transcripts, the system comprising a processing unit, the processing unit executing a software application configured for: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database.

According to a further aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts, the method comprising screening a database of expressed polynucleotides sequences according to at least one sequence criterion, the at least one sequence criterion being selected to identify putative naturally occurring antisense transcripts.

According to yet a further aspect of the present invention there is provided A method of quantifying at least one mRNA of interest in a biological sample, the method comprising: (a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one mRNA of interest, wherein the at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, the sequence region not being complementary with a naturally occurring antisense transcript; and (b) detecting a level of binding between the at least one mRNA of interest and the at least one oligonucleotide to thereby quantify the at least one mRNA of interest in the biological sample.

According to still a further aspect of the present invention there is provided a method of quantifying the expression potential of at least one mRNA of interest in a biological sample, the method comprising: (a) contacting

the biological sample with at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest; and (b) detecting a level of binding between the at least one mRNA of interest and the first oligonucleotide and a level of binding between the naturally occurring antisense transcript complementary to the mRNA of interest and the second oligonucleotide to thereby quantify the expression potential of the at least one mRNA of interest in the biological sample.

According to other aspect of the present invention there is provided a method of quantifying at least one naturally occurring antisense transcript of interest in a biological sample, the method comprising: (a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one naturally occurring antisense transcript of interest, wherein the at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the naturally occurring antisense transcript of interest, the sequence region not being complementary with a naturally occurring mRNA transcript; and (b) detecting a level of binding between the at least one naturally occurring antisense transcript of interest and the at least one oligonucleotide to thereby quantify the at least one naturally occurring antisense transcript of interest in the biological sample.

According to still further features in the described preferred embodiments the first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

According to still further features in the described preferred embodiments the second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA

(cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

According to still further features in the described preferred embodiments an average sequence length of the expressed polynucleotide sequences of the second database is selected from a range of 0.02 to 0.8 Kb.

According to still further features in the described preferred embodiments the second database is generated by: (i) providing a library of expressed polynucleotides; (ii) obtaining sequence information of the expressed polynucleotides; (iii) computationally selecting at least a portion of the expressed polynucleotides according to at least one sequence criterion; and (iv) storing the sequence information of the at least a portion of the expressed polynucleotides thereby generating the second database.

According to still further features in the described preferred embodiments the at least one sequence criterion for computationally selecting the at least a portion of the expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

According to still further features in the described preferred embodiments the step of testing the putative naturally occurring antisense transcripts for an ability to form the duplex with the at least one sense oriented polynucleotide sequence under physiological conditions.

According to still further features in the described preferred embodiments the method further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

According to still further features in the described preferred embodiments a length of the at least one oligonucleotide is selected from a range of 15-200 nucleotides.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a single stranded oligonucleotide.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a double stranded oligonucleotide.

According to still further features in the described preferred embodiments a guanidine and cytosine content of the at least one oligonucleotide is at least 25 %.

According to still further features in the described preferred embodiments the at least one oligonucleotide is labeled.

According to still further features in the described preferred embodiments the at least one oligonucleotide is attached to a solid substrate.

According to still further features in the described preferred embodiments the solid substrate is configured as a microarray and whereas the at least one oligonucleotide includes a plurality of oligonucleotides each attached to the microarray in a regio-specific manner.

According to still further features in the described preferred embodiments a length of each of the first and second oligonucleotides is selected from a range of 15-200 nucleotides.

According to still further features in the described preferred embodiments the first and second oligonucleotides are single stranded oligonucleotides.

According to still further features in the described preferred embodiments the first and second oligonucleotides are double stranded oligonucleotide.

According to still further features in the described preferred embodiments a guanidine and cytosine content of each of the first and second oligonucleotides is at least 25 %.

According to still further features in the described preferred embodiments the first and second oligonucleotides are labeled.

According to still further features in the described preferred embodiments the first and second oligonucleotides are attached to a solid substrate.

According to still further features in the described preferred embodiments the solid substrate is configured as a microarray and whereas each of the first and second oligonucleotides includes a plurality of oligonucleotides each attached to the microarray in a regio-specific manner.

According to yet other aspect of the present invention there is provided a method of identifying a novel drug target, the method comprising: (a) determining expression level of at least one naturally occurring antisense transcript of interest in cells characterized by an abnormal phenotype; and (b) comparing the expression level of the at least one naturally occurring antisense transcript of interest in the cells characterized by an abnormal phenotype to an expression level of the at least one naturally occurring antisense transcript of interest in cells characterized by a normal phenotype, to thereby identify the novel drug target.

According to still further features in the described preferred embodiments the abnormal phenotype of the cells is selected from the group consisting of biochemical phenotype, morphological phenotype and nutritional phenotype.

According to still further features in the described preferred embodiments determining expression level of at least one naturally occurring antisense transcript of interest is effected by at least one oligonucleotide designed and configured so as to be complementary to a sequence region of the

at least one naturally occurring antisense transcript of interest, the sequence region not being complementary with a naturally occurring mRNA transcript.

According to still other aspect of the present invention there is provided a method of treating or preventing a disease, condition or syndrome associated with an upregulation of a naturally occurring antisense transcript complementary to a naturally occurring mRNA transcript, the method comprising administering a therapeutically effective amount of an agent for regulating expression of the naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the agent for regulating expression of the naturally occurring antisense transcript is at least one oligonucleotide designed and configured so as to hybridize to a sequence region of the at least one naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a ribozyme.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a sense transcript.

According to a supplementary aspect of the present invention there is provided a method of diagnosing a disease, condition or syndrome associated with a substandard expression ratio of an mRNA of interest over a naturally occurring antisense transcript complementary to the mRNA of interest, the method comprising: (a) quantifying expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest; (b) calculating the expression ratio of the mRNA of interest over the naturally occurring antisense transcript complementary to the mRNA of interest, thereby diagnosing the disease, condition or syndrome.

The present invention successfully addresses the shortcomings of the presently known configurations by providing a novel approach for identifying naturally occurring antisense transcripts, methods of designing artificial

antisense transcripts according to information derived therefrom and methods and kits using naturally occurring and synthetic antisense transcripts.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIG. 1 illustrates EST alignment along genomic DNA, generated according to the teachings of the present invention. Alignment results identify two strand groups of transcripts i.e., sense transcripts and antisense transcripts with an indicated sequence overlap.

FIG. 2 illustrates a system designed and configured for generating a database of naturally occurring antisense sequences generated according to the teachings of the present invention.

FIG. 3 illustrates a remote configuration of the system described in Figure 2.

FIGs. 4a-k are sequence alignments of overlapping regions of selected naturally occurring antisense and sense sequence pairs identified according to the teachings of the present invention.

FIGs. 5a-g are sequence alignments of overlapping regions of selected naturally occurring antisense and sense sequence pairs identified according to the teachings of the present invention.

FIG. 6 schematically illustrates two transcription products of 53BP1 gene (red and green) and their corresponding partial complementary antisense transcripts of the 76p gene (blue). Numbers in parenthesis indicate length of sequence complementation. Schematic location of strand-specific RNA probes used for northern blotting of sense (53BP1, Riboprobe#1) and antisense (76p, Riboprobe#2) transcripts is shown.

FIG. 7 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of 53BP1 transcripts. Arrows on the right indicate the molecular weight of the identified 53BP1 transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. |Numbers on the left denote the size of molecular weight markers in Kb.

FIG. 8 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of 76p transcripts. Arrows on the right indicate the molecular weight of the identified 76p transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. |Numbers on the left denote the size of molecular weight markers in Kb.

FIG. 9 is an autoradiogram of a northern blot analysis depicting tissue distribution and expression levels of 76p transcripts. Arrows on the right indicate the molecular weight of the identified 76p transcripts. Numbers on the left denote the migration of molecular weight marker in Kb.

FIG. 10 illustrates the genomic organization of the 53BP1 gene and 76p gene, as elucidated from the RT-PCR analysis presented in the Examples section hereinbelow. Black arrows indicate the location of the primers used for RT-PCR analysis. Asterisks denote stop codons.

FIG. 11 schematically illustrates two transcription products of CIDE-B gene and their corresponding partial complementary antisense transcript of the BLTR2 gene. Schematic location of the strand-specific 430 nucleotide RNA

probe used for northern analysis of sense (CIDE-B) and antisense (BLTR2) transcripts is shown. Dashed rectangles indicate the predicted coding sequence of the transcripts.

FIG. 12 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of BLTR2 transcripts. Arrows on the right indicate the molecular weight of the identified BLTR2 transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. Numbers on the left denote the size of molecular weight markers in Kb.

FIG. 13 shows autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of CIDE-B transcripts. Arrows on the right indicate the molecular weight of the identified CIDE-B transcripts relatively to the migration of 28S and 18S ribosomal RNA subunits. Numbers on the left denote the migration size of molecular weight markers in Kb.

FIG. 14 schematically illustrates a transcription product of APAF-1 gene and its corresponding partial complementary antisense transcripts of the EB-1 gene. Schematic location of the strand-specific 366 nucleotide RNA probe used for northern analysis of sense (APAF-1) and antisense (EB-1) transcripts is shown. Asterisks indicate the predicted coding sequence borders of the transcripts.

FIGs. 15a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of EB-1 (Figure 15a) and APAF-1 transcripts (Figure 15b). Numbers on the left denote the size of molecular weight marker in Kb.

FIG. 16 schematically illustrates a transcription product of the MINK-2 gene and its corresponding partial complementary antisense transcript of the AchR-ε gene. Schematic location of the strand-specific 280 nucleotide RNA probe used for northern analysis of sense (Mink-2) and antisense (AchR-ε) transcripts is shown.

FIGs. 17a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of AchR-ε antisense transcripts

(Figure 17a) and the sense complementary transcript of Mink-2 (Figure 17b). Arrows on the right denote the migration of molecular weight markers in Kb.

FIG. 18 schematically illustrates a transcription product of Cyclin-E2 gene and its corresponding partial complementary antisense transcript. Schematic location of strand-specific RNA probes used for northern blotting of sense (Riboprobe#1) and antisense (Riboprobe#2) transcripts is shown.

FIGs. 19a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of Cyclin E2 antisense transcript (Figure 19a) and the sense complementary transcript (Figure 19b). Arrows on the left denote the migration of molecular weight markers in Kb.

FIG. 20 illustrates results from RT-PCR analysis of the expression patterns of CIDE-B transcript and its complementary naturally occurring antisense transcript following concentration dependent induction of apoptosis. Lanes: (1) 50 μM etoposide; (2) 100 μM etoposide; (3) 250 μM etoposide; (4) 500 μM etoposide; (5) 10 nM staurosporine; (6) 100 nM staurosporine; (7) 250 nM staurosporine; (8) 1000 nM staurosporine; (9) untreated cells (UT).

FIGs. 21a-c are results of RT-PCR analyses depicting expression patterns of AchRε and its naturally occurring antisense transcript following time-dependent induction of differentiation. Figure 21a illustrates the position of riboprobes used for reverse transcription reaction. Figure 21b shows the reciprocal expression pattern of sense and antisense transcripts (indicated by arrows). Figure 21c shows the expression pattern of the antisense transcript alone.

FIGs. 22a-j illustrate results of northern blot analysis of sense/antisense clusters revealing positive signals for sense/antisense genes in the microarray analysis. Diagrams describing genomic organization of the relevant region for each of the sense/antisense clusters are included above the autoradiograms, and regions of overlap (including GenBank accession number) from which the strand-specific riboprobes were derived are included. Sense-antisense pair numbers are as they appear in the microarray (as depicted in Table S2 on the

attached CD-ROM3 and in conversion Table 6). Figure 22a reveals expression patterns of randomly selected sequence pair number 235, denoted as Rand_235 in Table 6. Similarly, Figure 22b corresponds to pair number 173, Figure 22c to pair number 248, Figure 22d to pair number 6, Figure 22e to pair number 216, Figure 22f to pair number 239, Figure 22g to pair number 202, Figure 22h to pair number 114, Figure 22i to pair number 188, and Figure 22j to pair number 223. Eight pairs (Figures 22a-h) evaluated revealed positive signals for both sense and antisense expression, while two (Figures 22i-j) revealed a positive signal for only one of the genes, with the counterpart being a known RefSeq mRNA.

FIG. 23 is a Table depicting expression patterns in various cell lines and tissues as probed with a subset of 264 pairs from the putative sense/antisense dataset of the present invention. The pairs are denoted by the pair number and described in Table S1 of CD-ROM3. "C" and "AC" denote the two counterpart probes. Expression was also verified for positive controls, including the ubiquitously expressed genes *gapdh, actin, hsp70* and *gnb2l1* in various concentrations, and 11 previously documented sense/antisense pairs. Expression thresholds were verified and indicated as "+", if the probe passed the threshold in at least one cell line or tissue or "-", if the probe did not pass the threshold in all experiments. In cases where both the sense and the antisense oligo passed the expression threshold, the antisense was declared "verified". In cases where only one of the probes passed the expression threshold, but the other probe was fully contained within a known mRNA deposited in GenBank, the antisense was declared "indirectly verified". Normalization for microarray signals was conducted as described in the methods section. *Rji* ratios were obtained for each cell line/tissue assessed. Cases of flagged-out spots for which there was no information were marked "–1.00". Data represent values of the two reciprocal experiments.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of methods of identifying naturally occurring antisense transcripts, which can be used in kits and methods for quantifying gene expression levels. Specifically, the antisense molecules and related oligonucleotides generated according to information derived therefrom of the present invention can be used to detect, quantify, or specifically regulate antisense and respective sense transcripts thereby enabling detection and treatment of a wide range of disorders.

The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings described in the Examples section. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

Terminology

As used herein, the term "oligonucleotide" refers to a single stranded or double stranded oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics thereof. This term includes oligonucleotides composed of naturally-occurring bases, sugars and covalent internucleoside linkages (e.g., backbone) as well as oligonucleotides having non-naturally-occurring portions, which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases.

The term "antisense" refers to a complementary strand of an mRNA transcript e.g., antisense RNA.

The phrase "naturally occurring antisense transcripts" refers to RNA transcripts encoded from an antisense strand of the DNA. These endogenous transcript exhibit at least partial complementarity to mRNA transcripts transcribed from the sense strand of a DNA, also termed sense transcripts. *cis*-encoded naturally occurring antisense transcripts are transcribed from the same locus as the sense transcripts. *trans*-encoded antisense transcripts are transcribed from a different locus than the respective sense transcripts.

The phrase "antisense strand" or "anticoding strand" refers to a strand of DNA, which serves as a template for mRNA transcription and as such is complementary to the mRNA transcript formed.

The phrase "sense strand" or "coding strand" refers to the strand of DNA, which is identical to the mRNA transcript formed.

The phrase "complementary DNA" (cDNA) refers to the double stranded or single stranded DNA molecule, which is synthesized from a messenger RNA template.

The phrase "sense oriented polynucleotides" refers to polynucleotide sequences of a complementary or genomic DNA. Such polynucleotide sequences can be from exon regions, in which case they can encode mRNAs or portions thereof, or from intron regions, in which case they typically do not encode mRNA or portions thereof.

The term "contig" refers to a series of overlapping sequences with sufficient identity to create a longer contiguous sequence.

The term "cluster" refers to a plurality of contigs all derived, with a high degree of probability, from a single gene. Clusters are generally formed based upon a specified degree of homology and overlap (e.g., a stringency). The different contigs in a cluster do not typically represent the entire sequence of the gene, rather the gene may comprise one or more unknown intervening sequences between the defined contigs.

The phrase "open reading frame" (ORF) refers to a nucleotide sequence, which could potentially be translated into a polypeptide. Such a stretch of sequence is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons. For the purposes of this application, an ORF may be any part of a coding sequence, with or without start and/or stop codons. For an ORF to be considered as a good candidate for coding for a bona fide cellular protein, a minimum size requirement is often set, for example, a stretch of DNA that would code for a protein of 50 amino acids or more. An ORF is not usually considered an equivalent to a gene or locus until a phenotype is associated with a mutation in the ORF, an mRNA transcript for a gene product generated from the ORF's DNA has been detected, and/or the ORF's protein product has been identified.

The term "annotation" refers to a functional or structural description of a sequence, which may include identifying attributes such as locus name, poly(A)/poly(T) tail and/or signal, key words, Medline references and orientation cloning data.

Naturally occurring antisense molecules can play a role in sense transcription stability and function (e.g. translation). To date, most, if not all of the information relating to naturally occurring antisense transcripts was obtained by either low efficiency computational approaches (described hereinabove) or by approaches utilizing RNase protection assays, northern blot analysis, strand-specific RT PCR, subtractive hybridization, differential plaque hybridization, affinity chromatography, electrospray mass spectrometry and the like. These methods, though highly reliable, are extremely laborious, time consuming and are directed at individual target transcripts. As such, current approaches for uncovering antisense transcripts can be used to detect a negligible portion of the number of naturally occurring antisense molecules thought to exist.

As described hereinunder and in the Examples section, which follows, the present invention provides a novel approach for systematically identifying naturally occurring antisense molecules.

Aside from large scale applicability, the present method can be used to identify naturally occurring antisense molecules even in cases where the antisense transcriptional unit is localized to an intron of an expressed gene or to a different locus than the complementary sense encoding gene (e.g., trans-encoded antisense), or in cases where the antisense molecule lacks an open reading frame or appreciable complementarity to known sense molecules. Antisense transcripts uncovered according to the teachings of the present invention can be used for detecting and accurately quantifying respective sense counterparts as well as for sensibly designing artificial antisense molecules suitable for down-regulation of sense counterparts.

Thus, according to one aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts.

The method according to this aspect of the present invention is effected by the following steps.

First, sense-oriented polynucleotide sequences of a first database are computationally aligned with expressed polynucleotide sequences of a second database.

Following computational alignment, expressed polynucleotide sequences are analyzed according to one or more criteria for their ability to hybridize or form a duplex or partial complementation with the sense-oriented polynucleotide sequences (further detailed hereinbelow and in the Examples section which follows).

Expressed polynucleotide sequences which are capable of forming a duplex with sense oriented sequences are considered as putative naturally occurring antisense molecules and as such can be stored in a database which can be generated by a suitable computing platform.

Final confirmation of computationally obtained putative naturally occurring antisense molecules can be effected either computationally or preferably by using suitable laboratorial methodologies, based on nucleotide hybridization including RNase protection assay, subtractive hybridization, differential plaque hybridization, affinity chromatography, electrospray mass spectrometry, northern analysis, RT-PCR and the like (for further details see the Examples section).

Information derived from the sequence, sense position and other structure characteristics of the naturally occurring antisense transcripts identified according to the teachings of the present invention can be used to quantify respective sense transcripts of interest or to generate corresponding artificial antisense polynucleotides, which can be packed in diagnostic or therapeutic kits and implemented in various therapeutic and diagnostic methods.

Expressed polynucleotide sequences used as a potential source for identifying naturally occurring antisense transcripts according to this aspect of the present invention are preferably libraries of expressed messenger RNA [i.e., expressed sequence tags (EST), cDNA clones, contigs, pre-mRNA, etc.] obtained from tissue or cell-line preparations which can include genomic and/or cDNA sequence.

Expressed polynucleotide sequences, according to this aspect of the present invention can be retrieved from pre-existing publicly available databases (i.e., GenBank database maintained by the National Center for Biotechnology Information (NCBI), part of the National Library of Medicine, and the TIGR database maintained by The Institute for Genomic Research) or private databases (i.e., the LifeSeq.™ and PathoSeq.™ databases available from Incyte Pharmaceuticals, Inc. of Palo Alto, CA).

Alternatively, the sequence database of the expressed polynucleotide sequences utilized by the present invention can be generated from sequence

libraries (e.g., cDNA libraries, EST libraries, mRNA libraries and others). cDNA libraries are suitable sources for expressed sequence information.

Generating a sequence database in such a case is typically effected by tissue or cell sample preparation, RNA isolation, cDNA library construction and sequencing.

It will be appreciated that such cDNA libraries can be constructed from RNA isolated from whole organisms, tissues, tissue sections, or cell populations. Libraries can also be constructed from tissue reflecting a particular pathological or physiological state. Of particular interest are libraries constructed from sources associated with certain disease states, including malignant, neoplastic, hyperplastic tissues and the like.

Once raw sequence data is obtained, sequences are selected and preferably annotated before stored in a database. Selection proceeds according to one or more sequence criterion, which will be further detailed hereinunder. The editing, annotation and selection process is divided into two stages of processing. One stage comprises removal of repetitive, redundant or non-informative and contaminant sequences. The second stage involves selection of suitable candidates of putative naturally occurring antisense sequences.

The following section describes the different selection criteria which can be used for sequence filtering.

Vector contamination - "chops" vector elements and linker motifs used for the process of cloning from desired expressed nucleotide sequences. This selection can be effected by screening manually updated databases of sequences included in commonly used expression or cloning vectors.

Contaminating sequences - includes sequences which are derived from an undesired source. Such sequences can be recognized by their nucleotide distribution and/or by homology searches such as alignment searches using any sequence alignment algorithm such as BLAST (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST) or the Smith-Waterman algorithm. Other contaminating sequences may include sequences

exhibiting high occurrence of di-nucleotide distribution mostly related to sequencing artifacts and ribosomal RNA sequences.

Repetitive elements and low complexity sequences - eliminates or masks expressed sequences comprising known repetitive elements (ALU, L1 etc.) and low complexity sequences (i.e., a di- or tri-nucleotide repeat). Such elimination is preferably effected by comparison with database of known repetitive elements. It will be appreciated that this type of selection is mostly species specific. Masking of low complexity sequences can be effected by substituting an N (i.e., an inert character) for the actual nucleotide (i.e., G, A, T, or C). Masking of low complexity sequences facilitates further computational analysis and maintains the spacing of the molecule.

Sequence length - preferred expressed sequences are of a length between 20–2000, preferably 20-1000, more preferably 20-500, most preferably 20-300 base pairs.

Sequence annotation – expressed sequences retrieved from external databases, i.e., GenBank, oftentimes include an annotation which indicates direction of the sequencing of the insert clone (i.e., 5' or 3' direction). Sequence annotation, though "noisy" by nature due to multiple entries from various sources; artifacts taking place during directional cloning and incidence of palindromic eight-cutter restriction sites located at the end of the sequence, can serve as an important tool for deducing strand identity using dedicated computer software which are further discussed hereinunder

Intron splice site consensus sequence intron splice site sharing– intron sequences nearly always begin with a di-nucleotide sequence of GT ("splice donor") and end with an AG ("splice acceptor") preceded by a pyrimidine-rich tract. This consensus sequence is part of the signal for splicing. Intron splice site consensus sequence on the complementary strand (e.g., antisense strand) begins with CT and ends with AC. Thus, combined with genomic data, expressed sequences having a GT…AG can be considered as sense-oriented sequences, while a CT…AC pattern is considered as an antisense oriented

sequence. This selection criterion is very stringent since only negligible portions of introns have a CT...AC pattern. Sequences that share a similar splicing pattern, as deduced by alignment to genomic data, may be considered as having the same sense orientation, also termed herein as "intron sharing". It will be appreciated by one skilled in the art that using these selection criteria requires a careful and accurate alignment of expressed sequences to genomic sequence.

Poly(A) tails and Poly(T) heads – most eukaryotic mRNA molecules contain a poly-adenylation [poly(A)] tail at their 3' end. This poly(A) tail is not encoded by DNA. Therefore an expressed sequence which has a poly(A) tail can be considered as sense oriented. Similarly, poly(T) heads, which are not encoded from a genomic sequence indicate that a sequence is of the opposite direction, namely antisense oriented. Notably, genomically encoded Poly(A) tails and poly(T) heads provide no information as to the sequence orientation.

Poly(A) signal – some mature mRNA transcripts contain internal AAUAAA sequence. This internal sequence is part of an endonuclease cleavage signal. Following cleavage by the endonuclease, a poly(A) polymerase adds about 250 A residues to the 3' end of the transcript. Hence, expressed sequences containing a poly(A) signal can be considered as sense oriented.

Rare restriction site used for cloning– for example, eight cutter endonucleases which cleave 8-mer palindromic sequences and are characterized by a low frequency of cutting often used in genome mapping and EST library preparations (e.g., NotI. Commercially available from Promega: www.promega.com). Therefore, when a cluster of overlapping expressed sequences is characterized by a portion of sequences starting with a digestion site and another portion ending with the same, these sequences may be considered as encoded from the same strand. However, any endonuclease capable of digesting a palindromic sequence (i.e., XhoI, SalI, PacI etc.) may

also affect distorted sequence clustering, therefore strand orientation is preferably effected using other parameters as well.

Sequence overlap - sequences that completely overlap are considered to have the same strand orientation.

The above described parameters are used individually or in combination to analyze the expressed polynucleotide sequences so as to select anti-sense oriented sequences.

Selection can be effected on the basis of a single criterion or several criteria considered individually or in combination.

In cases where several criteria are examined, a scoring system e.g., a scoring matrix, is preferably used.

Since in some cases identifying an intron splicing consensus site may be more important than both sequence annotation and NotI alignment, while in others, detection of poly(A) tails and poly(T) heads might be the most significant criterion, the use of a scoring matrix in which each criterion is weighted enables one to select qualified antisense transcripts.

Such a scoring matrix can list the various expressed polynucleotide sequences across the X-axis of the matrix while each criterion can be listed on the Y-axis of the matrix. Criteria include both a predetermined range of values from which a single value is selected from each sequence, and a weight. Each sequence is scored at each criterion according to its value and the weight of the criterion.

When using such a scoring matrix the scores of each criterion of a specific sequence are summed and the results are analyzed.

Expressed sequences which exhibit a total score greater than a particular stringency threshold are grouped as members of either a sense-oriented sequence set or antisense-oriented sequence set; the higher the score the more stringent the criteria of grouping.

It will be appreciated that the above described analysis can take place prior to computational alignment to sense oriented sequences, i.e., during the

process of editing the expressed sequence database which is described hereinabove. Alternatively, selection can take place following computational alignment, thus further facilitating identification of proper duplex formation between the sense oriented polynucleotide sequences and expressed polynucleotide sequences.

Genomic DNA or a portion thereof is preferably used as sense-oriented sequence data according to this aspect of the present invention. It is conceivable that the present invention can determine sense orientation and antisense orientation of a database of expressed sequences simply by computationally aligning the sequences of the expressed database onto the genome, and finding whether two complementary expressed sequences hybridize to the genome (e.g., virtually generate a double stranded portion thereof). Such two overlapping sequences constitute sense and naturally occurring antisense transcripts.

Utilizing genomic DNA as a sense oriented template is preferred for the following reasons: (i) identifying trans-encoded antisense transcripts; (ii) analyzing intron splice consensus site and intron sharing; (iii) omitting genomically encoded poly(A) and poly(T) sequences; and (iv) analyzing sequences encompassing eight-cutter restriction sites.

Computational alignment of expressed polynucleotide sequences to the sense-oriented *polynucleotide sequences* (e.g., genomic sense sequences) can be effected using any commercially available alignment software, including sequence alignment tools utilizing algorithm such as BLAST (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST) or Smith-Waterman.

Assembly software is preferably used according to this aspect of the present invention. Such software is of high value when complete genomic information is unavailable or when handling large amounts of expressed sequence data. A number of commonly used computer software fragment read assemblers capable of forming clusters of expressed sequences are now

available. These packages include but are not limited to, The TIGR Assembler [Sutton G. et al. (1995) Genome Science and Technology 1:9-19], GAP [Bonfield JK. et al. (1995) Nucleic Acids Res. 23:4992-4999], CAP2 [Huang X. et al. (1996) Genomics 33:21-31], The Genome Construction Manager [Laurence CB. Et al. (1994) Genomics 23:192-201], Bio Image Sequence Assembly Manager, SeqMan [Swindell SR. and Plasterer JN. (1997) Methods Mol. Biol. 70:75-89], LEADS and GenCarta (Compugen Ltd. Israel).

Computer assembly and alignment programs can be modified to incorporate sequence criteria for determining sense or antisense orientation of expressed nucleotide sequences, as described hereinabove. Thereby, avoiding deliberate inversion of sequences during the assembly process, while ignoring the natural orientation of the sequences (i.e., sense or antisense orientation). Figure 1 illustrates results of expressed sequence assembly against genomic data and final distinction between sense oriented transcripts and antisense oriented transcripts of a single gene.

Following a proper alignment of expressed sequences to sense oriented polynucleotide sequences, duplexes are identified. The term "duplex" is used herein to indicate that a sequence identified according to this aspect of the present invention is complementary to a sense-oriented polynucleotide sequence. Complementation may be to a portion of the sense sequence, i.e., a region thereof, or alternatively, to two or more non-contiguous regions, which may be separated by one or more nucleotides on the sense strand.

The formation of sense-antisense duplexes does not require 100 % complementation nor does it require participation of the entire sense/antisense transcript sequence. The sense or antisense transcripts can have a secondary structure (e.g., stem and loop) generated by intra-sequence hybridization which can prevent specific sequence regions in the sense or antisense transcripts from participating in duplex formation. Thus, the antisense of the sequence identified, according to this aspect of the present invention can be

complementary to its sense counterparts in several regions, which are not necessarily close to each other when the sense transcript is in linear form.

Although any length of sequence overlap can generate a duplex, overlaps of at least 5 preferably 20 more preferably 30 even more preferably 40 bp are considered more indicative of true sense-antisense duplex formation.

The method of uncovering putative antisense transcripts of the present invention is preferably carried out using a dedicated computational system.

Thus, according to another aspect of the present invention and as illustrated in Figure 2, there is provided a system for generating a database of putative naturally occurring antisense sequences which system is referred to hereinunder as system **10**.

System **10** includes a processing unit **12**, which executes a software application designed and configured for aligning sense oriented polynucleotide sequences with expressed polynucleotide sequences and identifying expressed polynucleotide sequences which are capable of forming a duplex with the sense oriented polynucleotide sequences, thereby recognizing putative naturally occurring antisense transcripts. System **10** may also include a user input interface **14** (e.g., a keyboard and/or a mouse) for inputting database or database related information, and a user output interface **16** (e.g., a monitor) for providing database information to a user.

System **10** preferably stores sequence information of the putative antisense transcripts identified thereby on a computer readable media such as a magnetic, optico-magnetic or optical disk to thereby generate a database of putative antisense transcript sequences. Such a database further includes information pertaining to database generation (e.g., source library), parameters used for selecting polynucleotide sequences, putative uses of the stored sequences, and various other annotations and references which relate to the stored sequences or respective sense transcripts.

System **10** of the present invention may be used by a user to query the stored database of sequences, to retrieve nucleotide sequences stored therein or to generate polynucleotide sequences from user inputted sequences.

System **10** can be any computing platform known in the art including but not limited to, a personal computer, a work station, a mainframe and the like.

The database generated and stored by system **10** can be accessed by an on-site user of system **10**, or by a remote user communicating with system **10**.

As illustrated in Figure 3, communication between a remote user **18** and processing unit **12** is preferably effected via a communication network **20**. Communication network **20** can be any private or public communication network including, but not limited to, a standard or cellular telephony network, a computer network such as the Internet or intranet, a satellite network or any combination thereof.

As illustrated in Figure 3, communication network **20** includes one or more communication servers **22** (one shown in Figure 3) which serves for communicating data pertaining to the polypeptide of interest between remote user **18** and processing unit **12**.

It will be appreciated that existing computer networks such as the Internet can provide the infrastructure and technology necessary for supporting data communication between any number of sites **24** and remote analysis sites **26**.

For example, using a computer operating a Web browser application and the World Wide Web, any expressed polynucleotide sequence of interest can be "uploaded" by user **18** onto a Web site maintained by a database server **28**. Following uploading, database server **28** which serves as processing unit **12** can be instructed by the user to processes the polynucleotide as is described hereinabove.

Following such processing, which can be performed in real time, nucleic acid sequence results can be displayed at the web site maintained by database server **28** and/or communicated back to site **24**, via for example, e-mail communication.

Thus, using the Internet, a remote configuration of system **10** can provide polynucleotide sequence analysis services to a plurality of sites **24** (one shown in Figure 3).

It will be appreciated that this configuration of system **10** of the present invention is especially advantageous in cases where polypeptide analysis can not be effected on-site. For example, laboratories, which lack the equipment necessary for executing the analysis or lack the necessary skills to operate it.

Thus, data extracted from the database of naturally occurring antisense transcripts of the present invention is of high value for designing oligonucleotides suitable for transcript detection and quantification and for sensibly designing artificial antisense oligonucleotides for down-regulation and elimination of a transcript of interest or changing the balance between sense and complementary antisense transcripts. The possibility of up-regulating a transcript of interest using naturally occurring antisense based-oligonucleotides generated according to the teachings of the present invention is also realized. In addition, data extracted from the database of naturally occurring antisense transcripts may also be used for assessing endogenous double stranded-RNA also termed interfering RNA, which may distort gene-expression due to either RNA-degradation, DNA-methylation, polycomb mediated suppression etc. (for details see the Background section hereinabove).

Antisense technology is based upon the pairing of an artificially designed antisense oligonucleotide, with a target nucleic acid. The use of antisense technology requires a complementarity of the antisense nucleotide sequence to a target zone of an mRNA target sequence that will effect inhibition of gene expression [reviewed in Stein CA. and Cohen JS. (1988) Cancer Res. 48:2659-68]. Based on empiric experience it was shown that the

success of antisense technology relies on: (i) cellular uptake; (ii) stability of artificial antisense molecules under physiological conditions (i.e., cellular pH, endonucleases etc.); (iii) complementation between the oligonucleotide and a single stranded target sequence (i.e., tertiary structure of target RNA will not form a good target); (iv) binding specificity of antisense oligonucleotide so as not to compete with other RNA binders (e.g. proteins) to thereby maintain an effective antisense concentration.

Various attempts to employ antisense technology while considering the above discussed limitations included using large amounts of oligonucleotides to overcome cellular uptake and environmental barriers and chemically modified antisense nucleotide compositions, for obtaining higher level of cellular stability. However, even in case where uptake difficulties are traversed, the step of target identification (i.e., RNA-target sequence region) continues to be the major bottleneck for successful implementation of antisense technology.

U.S. Pat. No: 6,183,966 discloses a method and an apparatus for ranking nucleic acid sequences based on stability of nucleic acid oligomer sequence binding interactions to select sequence zones for antisense targeting. This method however systematic, relies on thermodynamic analyses combined with numerous predictions which cannot be considered empirically accurate and reliable.

Thus according to another aspect of the present invention there is provided a method of designing artificial antisense transcripts.

The method according to this aspect of the present invention is effected by the following steps.

First, structural and/or functional parameters pertaining to naturally occurring antisense transcripts are extracted/deduced from a database such as the one described hereinabove. These parameters may be generally deduced from all sequences stored in the database, or extracted from specific antisense sequences or preferably groups of antisense sequences.

Second, artificial antisense molecules of interest are designed according to the extracted parameters.

Such parameters may be divided into three groups, topographical parameters, functional parameters and structural parameters.

Topographical parameters - (i) position of sequence overlap on the sense transcript (i.e., coding region, 5'UTR, 3'UTR); (ii) position of the sequence overlap on the antisense transcript (end overlap, middle overlap, full overlap). (iii) length of overall sequence overlap; (iv) continuity or discontinuity of sequence overlap.

Structural parameters - pertains to both sense and antisense transcripts (i) tertiary structure (i.e., hairpin, helix, stem and loop, pseudoknot, and the like); (ii) single stranded versus double stranded regions; (iii) GC content; (iv) tandem Gs; (v) adenosine/inosine content; (vi) thermodynamic stability of tertiary structures; (vii) duplex melting point; (viii) methylations and other RNA modifications; (ix) RNA-protein interactions ; and (x) transcript length.

Functional parameters - (i) alternative splicing; (ii) tissue expression; (iii) pathology specific expression; (iv) antisense promoters; (v) intron content; (vi) open reading frame in antisense transcript.

These parameters can be used individually or in combination, in which case, each parameter is preferably weighted according to its importance. Due to the multi-factorial design of artificial antisense transcripts according to this aspect of the present invention, employing a scoring system (described hereinabove) is preferably used to simplify and increase the accuracy of the process.

Synthetic antisense oligonucleotides designed according to the teachings of the present invention can be generated according to any oligonucleotide synthesis method known in the art such as enzymatic synthesis or solid phase synthesis. Equipment and reagents for executing solid-phase synthesis are commercially available from, for example, Applied Biosystems. Any other

means for such synthesis may also be employed; the actual synthesis of the oligonucleotides is well within the capabilities of one skilled in the art.

Oligonucleotides used according to this aspect of the present invention are those having a length selected from a range of 10 to about 200 bases preferably 15-150 bases, more preferably 20-100 bases, most preferably 20-50 bases.

The oligonucleotides of the present invention may comprise heterocylic nucleosides consisting of purines and the pyrimidines bases, bonded in a 3' to 5' phosphodiester linkage.

Preferably used oligonucleotides are those modified in either backbone, internucleoside linkages or bases, as is broadly described hereinunder. Such modifications can oftentimes facilitate oligonucleotide uptake and resistance to intracellular conditions.

Specific examples of preferred oligonucleotides useful according to this aspect of the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone, as disclosed in U.S. Pat. NOs: ,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466, 677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkyl phosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having

inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms can also be used.

Alternatively, modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH2 component parts, as disclosed in U.S. Pat. Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623, 070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439.

Other oligonucleotides which can be used according to the present invention, are those modified in both sugar and the internucleoside linkage, i.e., the backbone, of the nucleotide units are replaced with novel groups. The base units are maintained for complementation with the appropriate polynucleotide target. An example for such an oligonucleotide mimetic, includes peptide nucleic acid (PNA). A PNA oligonucleotide refers to an oligonucleotide where the sugar-backbone is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The bases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,082; 5,714,331; and

5,719,262, each of which is herein incorporated by reference. Other backbone modifications, which can be used in the present invention are disclosed in U.S. Pat. No: 6,303,374. Oligonucleotides of the present invention may also include base modifications or substitutions. As used herein, "unmodified" or "natural" bases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified bases include but are not limited to other synthetic and natural bases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further bases include those disclosed in U.S. Pat. No: 3,687,808, those disclosed in The Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those disclosed by Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613, and those disclosed by Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B. , ed., CRC Press, 1993. Such bases are particularly useful for increasing the binding affinity of the oligomeric compounds of the invention. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 °C. [Sanghvi YS et al. (1993) Antisense Research and Applications, CRC Press, Boca Raton 276-278] and are presently

preferred base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more moieties or conjugates, which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-S-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety, as disclosed in U.S. Pat. No: 6,303,374.

It is not necessary for all positions in a given oligonucleotide molecule to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide.

The present invention also includes antisense molecules, which are chimeric molecules. "Chimeric" antisense molecules", are oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one nucleotide. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target polynucleotide. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. An example for such include RNase H, which is a cellular endonuclease which cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable

results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Cleavage of the RNA target can be routinely detected by gel electrophoresis and, if necessary, associated nucleic acid hybridization techniques known in the art.

Chimeric antisense molecules of the present invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, as described above. Representative U.S. patents that teach the preparation of such hybrid structures include, but are not limited to, U.S. Pat. Nos. 5,013,830; 5,149,797; 5,220,007; 5,256,775; 5,366,878; 5,403,711; 5,491,133; 5,565,350; 5,623,065; 5,652,355; 5,652,356; and 5,700,922, each of which is herein fully incorporated by reference.

Finally, chimeric oligonucleotides of the present invention can comprise a ribozyme sequence. Ribozymes are being increasingly used for the sequence-specific inhibition of gene expression by the cleavage of mRNAs. Several ribozyme sequences can be fused to the oligonucleotides of the present invention. These sequences include but are not limited ANGIOZYME specifically inhibiting formation of the VEGF-R (Vascular Endothelial Growth Factor receptor), a key component in the angiogenesis pathway, and HEPTAZYME, a ribozyme designed to selectively destroy Hepatitis C Virus (HCV) RNA, (Ribozyme Pharmaceuticals, Incorporated - WEB home page).

The oligonucleotides generated according to the teachings of the present invention can be used for both diagnostic and therapeutic purposes. For example, oligonucleotides of the present invention can be used to diagnose and treat a variety of diseases or pathological conditions associated with an abnormal expression (i.e., up-regulation or down-regulation) of at least one mRNA molecule of interest, including but not limited to diabetes, autoimmune diseases, Parkinson, Alzheimer' disease, HIV, malaria, cholera, influenza, rabies, diphtheria, breast cancer, colon cancer, cervical cancer, melanoma, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, lymphomas,

leukemias and the like and any other diseases (see Example 8 of the Examples section) which are associated with aberrant expression of multiple mRNAs (i.e., sense and/or antisense) or with unregulated formation of endogenous double stranded RNA complexes.

Present-day mRNA-based diagnostic assays utilize oligonucleotide probes which are complementary to one or more regions of the mRNA to be quantitated. Such probes are designed while considering interspecies sequence variation, sequence length, GC content etc. However design of such prior art probes (i.e., riboprobes or deoxyriboprobes) does not take into consideration the presence of antisense transcripts which can effect probe binding efficiency. Discounting antisense presence can lead to inaccurate diagnosis, which is oftentimes followed by an erroneous treatment protocol.

The present invention provides an mRNA-detection/quantification assay, which is devoid of this limitation.

Thus, according to an additional aspect of the present invention there is provided a method of quantifying at least one mRNA of interest in a biological sample.

As used herein, the phrase "biological sample" refers to any sample derived from biological tissues or fluids, including blood (serum or plasma), sputum, pleural effusions, urine, biopsy specimens, isolated cells and/or cell membrane preparation. Methods of obtaining tissue biopsies and body fluids from mammals are well known in the art.

The method of this aspect of the present invention is effected by contacting mRNA from a cell type or within a cell with one or more oligonucleotides that hybridizes efficiently with a sequence region of an mRNA transcript which is not complementary with a naturally occurring antisense transcript.

In addition to the limitation described above, prior art diagnostic/detection assays also fail to consider the effect of antisense transcription on the protein expression levels of a gene of interest. It stands to

reason that presence of antisense transcripts in a biological sample can substantially reduce the resultant protein levels translated from a complementary sense transcript. Consistently, diseases which are associated with endogenous dsRNA complexes, are also very difficult to detect and moreover to treat, due to insufficient sequence data pertaining to duplex forming transcripts.

Thus, for accurate quantification of gene expression, both the sense and antisense levels must be quantified and/or their respective expression ratio must be determined.

By contacting a biological sample with one or more pairs of oligonucleotides, where one oligonucleotide is capable of hybridizing with the mRNA of interest and the second oligonucleotide is capable of hybridizing with a naturally occurring antisense transcript which is complementary with the mRNA of interest such accurate quantification can be effected.

Contacting the oligonucleotides of the present invention with the biological sample is effected by stringent, moderate or mild hybridization (as used in any polynucleotide hybridization assay such as northern blot, dot blot, RNase protection assay, RT-PCR and the like). Wherein stringent hybridization is effected by a hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 1 - 1.5 °C below the Tm, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the Tm; moderate hybridization is effected by a hybridization solution of 6 x SSC and 0.1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 2 - 2.5 °C below the Tm, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the Tm, final wash solution of 6 x SSC, and final wash at 22 °C;

whereas mild hybridization is effected by a hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 37 °C, final wash solution of 6 x SSC and final wash at 22 °C.

The oligonucleotides of the present invention can be attached to a solid substrate, which may consist of a particulate solid phase such as nylon filters, glass slides or silicon chips [Schena et al. (1995) Science 270:467-470].

In a particular embodiment, oligonucleotides of the present invention can be attached to a solid substrate, which is designed as a microarray. Microarrays are known in the art and consist of a surface to which probes that correspond in sequence to gene products (e.g., cDNAs, mRNAs, cRNAs, polypeptides, and fragments thereof), can be specifically hybridized or bound at a known position (regiospecificity).

Several methods for attaching the oligonucleotides to a microarray are known in the art including but not limited to glass-printing, described generally by Schena et al., 1995, Science 270:467-47, photolithographic techniques [Fodor et al. (1991) Science 251:767-773], inkjet printing, masking and the like.

In general, quantifying hybridization complexes is well known in the art and may be achieved by any one of several approaches. These approaches are generally based on the detection of a label or marker, such as any radioactive, fluorescent, biological or enzymatic tags or labels of standard use in the art. A label can be applied on either the oligonucleotide probes or nucleic acids derived from the biological sample.

The following illustrates a number of labeling methods suitable for use in the present invention. For example, oligonucleotides of the present invention can be labeled subsequent to synthesis, by incorporating biotinylated dNTPs or rNTP, or some similar means (e.g., photo-cross-linking a psoralen derivative of biotin to RNAs), followed by addition of labeled streptavidin (e.g.,

phycoerythrin-conjugated streptavidin) or the equivalent. Alternatively, when fluorescently-labeled oligonucleotide probes are used, fluorescein, lissamine, phycoerythrin, rhodamine (Perkin Elmer Cetus), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX (Amersham) and others [e.g., Kricka et al. (1992), Academic Press San Diego, Calif] can be attached to the oligonucleotides. It will be appreciated that pairs of fluorophores are chosen when distinction between two emission spectra of two oligonucleotides is desired or optionally, a label other than a fluorescent label is used. For example, a radioactive label, or a pair of radioactive labels with distinct emission spectra, can be used [Zhao et al. (1995) Gene 156:207]. However, because of scattering of radioactive particles, and the consequent requirement for widely spaced binding sites, the use of fluorophores rather than radioisotopes is more preferred.

The intensity of signal produced in any of the detection methods described hereinabove may be analyzed manually or using a software application and hardware suited for such purposes.

In general, mRNA quantification is preferably effected alongside a calibration curve so as to enable accurate mRNA determination. Furthermore, quantifying transcript(s) originating from a biological sample is preferably effected by comparison to a normal sample, which sample is characterized by normal expression pattern of the examined transcript(s).

It will be appreciated that the detection method described above can also be used for quantifying at least one naturally occurring antisense transcript in a biological sample. In such a case, the oligonucleotide used for quantification is designed to hybridize with a sequence region of naturally occurring antisense transcript of interest, which is not complementary with a naturally occurring mRNA transcript.

The diagnostic assays described hereinabove can be used to accurately distinguish between absence, presence and excess expression of any transcripts of interest (e.g., sense, antisense), and to monitor their level during therapeutic intervention. These methods are also capable of diagnosing diseases associated

with an improper balance or ratio between sense and antisense expression and diseases associated with endogenous dsRNA.

Further description of oligonucleotide-pair arrays is provided in Example 9 of the Examples section which follows.

As discussed hereinabove oligonucleotides of the present invention can be also used for therapeutic purposes, such as treating diseases or conditions associated with aberrant expression levels of one or more sense and/or antisense transcripts and conditions, which are associated with endogenous dsRNA such as unregulated formation of double-strand RNA (i.e., up/down-regulation).

Accumulative knowledge shows strong correlation between a variety of human diseases and mutations, over-expression and function of the protein building blocks (i.e., protein kinases, phosphatsases) and their effectors and regulators, which constitute numerous intracellular signaling pathways. For instance, inactivation of both copies of ZAP-70 or Jak-3 causes severe combined immunodeficiency and mutation of the X-linked BTK gene results in agammaglobulinemia. Many genetic disorders are also associated with mutations for example, in protein-serine kinases (PSKs) and phosphatases. The Coffin-Lowry syndrome results from inactivation of the X-linked Rsk2 gene, and myotonic dystrophy is due to decreased levels of expression of the myotonic dystrophy PSK. In addition, over-expression of ErbB2 receptor tyrosine kinase is implicated in breast and ovarian carcinoma [reviewed by Hunter T. (2000) Cell 100:113-127] .

Given the importance of activated kinases in a variety of disorders such as cancer, it would be anticipated that phosphatases regulation would be found as tumor suppressor genes and as promising drug targets. So far this has not proved to be the case. Furthermore, a number of diseases are associated with insufficient expression of signaling molecules, including non-insulin-dependent diabetes and peripheral neuropathies.

Thus, it is conceivable that identification of naturally occurring antisense transcripts of signaling molecules participating in specified signaling pathways may serve as promising tools for both identification and particularly treatment of a variety of disorders at any gene expression level (i.e., RNA, DNA or protein).

The term "treating" refers to alleviating or diminishing a symptom associated with the disease or the condition. Preferably, treating cures, e.g., substantially eliminates, and/or substantially decreases, the symptoms associated with the diseases or conditions of the present invention.

The treatment method according to the teachings of the present invention includes administering to an individual a therapeutically effective amount of the synthetic antisense oligonucleotides of the present invention. Preferred individual subjects according to the present invention are mammals such as canines, felines, ovines, porcines, equines, bovines, humans and the like.

A therapeutically effective amount implies an amount of agent effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the individual being treated

The agent of the method of the present invention can be administered to an individual *per se*, or as part of a pharmaceutical composition where it is mixed with a pharmaceutically acceptable carrier.

As used herein a "pharmaceutical composition" refers to a composition of one or more of the agents described hereinabove, or physiologically acceptable salts or prodrugs thereof, with other chemical components. The purpose of a pharmaceutical composition is to facilitate administration of a compound to an organism.

The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including vaginal and rectal delivery), pulmonary, e.g., by inhalation or insufflation of powders or

aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration. Oligonucleotides with at least one 2'-O-methoxyethyl modification are believed to be particularly useful for oral administration.

Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Coated condoms, gloves and the like may also be useful.

Compositions and formulations for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. Thickeners, flavoring agents, diluents, emulsifiers, dispersing aids or binders may be desirable.

Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, and liposome-containing formulations. These compositions may be generated from a variety of components that include, but are not limited to, preformed liquids, self-emulsifying solids and self-emulsifying semisolids.

The pharmaceutical formulations of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the formulations

are prepared by uniformly and intimately bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, liquid syrups, soft gels, suppositories, and enemas. The compositions of the present invention may also be formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

In one embodiment of the present invention the pharmaceutical compositions may be formulated and used as foams. Pharmaceutical foams include formulations such as, but not limited to, emulsions, microemulsions, creams, jellies and liposomes. While basically similar in nature these formulations vary in the components and the consistency of the final product. The preparation of such compositions and formulations is generally known to those skilled in the pharmaceutical and formulation arts and may be applied to the formulation of the compositions of the present invention.

The pharmaceutical compositions of the present invention may employ various penetration enhancers to effect the efficient delivery of nucleic acids, particularly oligonucleotides, to the skin of animals.

Penetration enhancers may be classified as belonging to one of five broad categories, i.e., surfactants, fatty acids, bile salts, chelating agents, and non-chelating non-surfactants [Lee et al., Critical Reviews in Therapeutic Drug Carrier Systems (1991) 92] as disclosed in U.S. Pat. No: 6,300,132, 6,271,030, 6,277,633, 6,284,538, 6,287,860, 6,294,382, 6,277,640 and 6,258,601 each of which is herein fully incorporated by reference.

Other substances that enhance uptake of oligonucleotides at the cellular level may also be added to the pharmaceutical compositions of the present

invention. For example, cationic lipids, such as lipofectin [U.S. Pat. No. 5,705,188], cationic glycerol derivatives, and polycationic molecules, such as polylysine [PCT Application WO 97/30731], are also known to enhance the cellular uptake of oligonucleotides.

Other reagents may be utilized to enhance the penetration of the administered nucleic acids, including glycols such as ethylene glycol and propylene glycol, pyrrols such as 2-pyrrol, azones, and terpenes such as limonene and menthone.

Certain pharmaceutical compositions of the present invention may also incorporate carrier compounds. As used herein, "carrier compound" or "carrier" can refer to a nucleic acid, or analog thereof, which is inert (i.e., does not possess biological activity per se) but is recognized as a nucleic acid by in vivo processes that reduce the bioavailability of a nucleic acid having biological activity by, for example, degrading the biologically active nucleic acid or promoting its removal from circulation. The co-administration of a nucleic acid and a carrier compound, typically with an excess of the latter substance, can result in a substantial reduction of the amount of nucleic acid recovered in the liver, kidney or other extracirculatory reservoirs, presumably due to competition between the carrier compound and the nucleic acid for a common receptor. For example, the recovery of a partially phosphorothioate oligonucleotide in hepatic tissue can be reduced when it is coadministered with polyinosinic acid, dextran sulfate, polycytidic acid or 4-acetamido-4' isothiocyano-stilbene-2,2'-disulfonic acid [Miyao et al., Antisense Res. Dev., (1995) 5:115-121; Takakura et al., Antisense & Nucl. Acid Drug Dev. (1996) 6:177-183].

In contrast to a carrier compound, an "excipient" is a pharmaceutically acceptable solvent, suspending agent or any other pharmacologically inert vehicle for delivering one or more nucleic acids to an animal. The excipient may be liquid or solid and is selected, with the planned manner of administration in mind, so as to provide for the desired bulk, consistency, etc., when combined with a nucleic acid and the other components of a given

pharmaceutical composition. Typical excipients include, but are not limited to, binding agents (e.g., pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose, etc.); fillers (e.g., lactose and other sugars, microcrystalline cellulose, pectin, gelatin, calcium sulfate, ethyl cellulose, polyacrylates or calcium hydrogen phosphate, etc.); lubricants (e.g., magnesium stearate, talc, silica, colloidal silicon dioxide, stearic acid, metallic stearates, hydrogenated vegetable oils, corn starch, polyethylene glycols, sodium benzoate, sodium acetate, etc.); disintegrants (e.g., starch, sodium starch glycolate, etc.); and wetting agents (e.g., sodium lauryl sulphate, etc.).

Pharmaceutically acceptable organic or inorganic excipient suitable for non-parenteral administration which do not deleteriously react with nucleic acids can also be used to formulate the compositions of the present invention. Suitable pharmaceutically acceptable carriers include, but are not limited to, water, salt solutions, alcohols, polyethylene glycols, gelatin, lactose, amylose, magnesium stearate, talc, silicic acid, viscous paraffin, hydroxymethylcellulose, polyvinylpyrrolidone and the like.

Formulations for topical administration of nucleic acids may include sterile and non-sterile aqueous solutions, non-aqueous solutions in common solvents such as alcohols, or solutions of the nucleic acids in liquid or solid oil bases. The solutions may also contain buffers, diluents and other suitable additives. Pharmaceutically acceptable organic or inorganic excipients suitable for non-parenteral administration, which do not deleteriously react with nucleic acids can be used.

Suitable pharmaceutically acceptable excipients include, but are not limited to, water, salt solutions, alcohol, polyethylene glycols, gelatin, lactose, amylose, magnesium stearate, talc, silicic acid, viscous paraffin, hydroxymethylcellulose, polyvinylpyrrolidone and the like.

The compositions of the present invention may additionally contain other adjunct components conventionally found in pharmaceutical compositions, at their art-established usage levels. Thus, for example, the compositions may

contain additional, compatible, pharmaceutically-active materials such as, for example, antipruritics, astringents, local anesthetics or anti-inflammatory agents, or may contain additional materials useful in physically formulating various dosage forms of the compositions of the present invention, such as dyes, flavoring agents, preservatives, antioxidants, opacifiers, thickening agents and stabilizers. However, such materials, when added, should not unduly interfere with the biological activities of the components of the compositions of the present invention. The formulations can be sterilized and, if desired, mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, colorings, flavorings and/or aromatic substances and the like which do not deleteriously interact with the nucleic acid(s) of the formulation. Aqueous suspensions may contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

The formulation of therapeutic compositions and their subsequent administration is believed to be within the skill of those in the art. Dosing is dependent on severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on EC50 found to be effective in in vitro and in vivo animal models. Persons of ordinary skill in the art can easily estimate dosing and repetition rates based on measured residence times and concentrations of the oligonucleotide in bodily fluids or tissues. Following successful treatment, it may be desirable to have the patient undergo

maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is administered in maintenance doses.

The methods of the present invention have evident utility in the diagnosis and treatment of various diseases and conditions. In addition, such methods can also be used in non-clinical applications, such as, for example, differential cloning, detection of rearrangements in DNA sequences as disclosed in U.S. Pat. No: 5,994,320, drug discovery and the like.

The oligonucleotides generated according to the teachings of the present invention can be included in a diagnostic or therapeutic kit. For example, oligonucleotides sets pertaining to specific disease related transcripts can be packaged in a one or more containers with appropriate buffers and preservatives along with suitable instructions for use and used for diagnosis or for directing therapeutic treatment.

Preferably, the containers include a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers may be formed from a variety of materials such as glass or plastic.

In addition, other additives such as stabilizers, buffers, blockers and the like may also be added.

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

EXAMPLES

Reference is now made to the following examples, which together with the above descriptions, illustrate the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical,

microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak et al., "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are

believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

In-vitro expression substantiation of computationally retrieved naturally occurring antisense transcripts

In-vitro expression assays were conducted in order to validate the existence of naturally occurring antisense sequences identified according to the teachings of the present invention.

Table 1 below lists polynucleotide sequence pairs that were selected for the *in-vitro* expression validation assays described in examples 1-7.

Table 1

Name of sense antisense pair	Sense transcript	Sense Length (nt)	Antisense transcript	Anti-sense Length (nt)	Overlap length (nt)	Start of overlap sense transcript	Start of overlap anti-sense
53BP1_76P	53BP1	10394	76P	6837	3046	5463	2018
	(SEQ ID NO: 15)		(SEQ ID NO: 16)				
CIDEB_BLTR2 (1)	CIDEB1	2289	BLTR2	6530	2254	17	1
	(SEQ ID NO: 19)		(SEQ ID NO: 21)				
CIDEB_BLTR2 (2)	CIDEB2	1511	BLTR2	6530	1410	1	1
	(SEQ ID NO: 20)						
APAF1_EB1	aAPAF1	7042	EB1a	1752	141	6889	1612
	(SEQ ID NO: 24)		(SEQ ID NO: 25)				
AChR_MINK2	AchR	2457	MINK2	4863	236	2175	4853
	(SEQ ID NO: 29)		(SEQ ID NO: 30)				
M-AchR_Anti-AChR	M-AchR	1590	M-Anti-AchR	2227	672	934	506
	(SEQ ID NO: 35)		(SEQ ID NO: 36)				
CyclinE2_Anti-CyclinE2	CyclinE2	2714	Anti-CyclinE2	5773	1855	565	2006
	(SEQ ID NO: 33)		(SEQ ID NO: 34)				

Sequence alignments of overlapping regions of each sense-antisense pair were performed using the BLAST sequence alignment algorithm (Basic Local

Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST using the default parameters) and are exhibited in Figure 5a-g.

A microarray-based analysis was conducted, as well, in order to validate the existence of naturally occurring, antisense sequences identified according to the teachings of the present invention. The results are described in Example 9.

Materials and Experimental Methods

RNA probes generation and northern analysis

RNA probes for northern analysis were generated by PCR amplification of a desired DNA fragment and cloning into Zero Blunt TOPO (Invitrogen Corp.) or pSPT18/19 vectors (Roche Ltd.). Alternatively PCR products were ligated into T7 RNA polymerase promoter-containing adaptors using the Lignscribe kit (Ambion Europe Ltd.). Corresponding RNA transcripts were synthesized using T7 RNA polymerase (Roche Ltd.) and labeled with 32P-UTP according to manufacturer's instructions. RNA probes were purified on Mini Quick Spin RNA columns.

Commercial membranes containing Poly(A)-RNA from various human tissues (2 μg RNA per lane) were obtained from Origene (OriGene Technologies Inc.) and Ambion (Ambion Inc.).

Alternatively, 2 μg of poly(A)-RNA prepared from various human cell-lines were electrophoretically separated on 1 % agarose gel, and electrotransferred to Nytran SuperCharge membrane (Schleicher & Schuell) and subjected to fixing by UV radiation. Membranes were stained with methylene blue to ensure quantitative RNA transfer. Membranes were then prehybridized in a hybridization solution (UltraHyb solution Ambion Europe Ltd.) for 30 minutes at 68 °C in a rotating hybridization tube.

Hybridization solution was then supplemented with 106 cpm of labeled RNA probe per each ml of hybridization solution. Blots were hybridized for 16 hours at 68 °C in a rotating hybridization tube. Membranes were then washed twice with 2 x SSC, 0.1 % sodium dodecyl sulfate (SDS) and twice with 0.1 %

SDS at 68 °C. RNA transcripts signals were detected using a phosphoimager (Molecular Dynamics, Sunnyvale CA).

Microarray

Oligonucleotide design - oligonucleotide design tools (1) were applied to each pair of sense/antisense genes in order to select two complementary 60-mer oligonucleotides from the region where the two genes overlap. The design criteria included the following: low cross-homology (up to 75%) to other expressed sequences in the human transcriptome; a continuous hit of no more than 17 bp to the sequence of another gene; balanced GC content (30-70%) without significant windows of local imbalance; no more than 2 palindromes with a length of 6 bp; a hit of no more than 15 bp to a repeat, vector or low-complexity region; and no long stretches of identical nucleotides.

Microarray preparation - 60-mer oligonucleotides were synthesized by Sigma-Genosys (The Woodlands, TX), resuspended at 40 μM in 3X SSC, and spotted in quadruplicates on poly-L-lysine coated glass slides as detailed in the online protocol of the National Human Genome Research Institute (http://www.nhgri.nih.gov/DIR/Microarray/Protocols.pdf). To avoid local differences in the hybridization conditions, the probes selected from the overlapping regions of each sense/antisense pair were spotted in the same block, next to each other.

Human cell lines - The following cell lines utilized were purchased from ATCC (Manassas, VA): MCF7 (breast adenocarcinoma, Cat. No. HTB-22,), HeLa (cervical adenocarcinoma, Cat. No. CCL-2) HEK-293 (embryonal kidney cells, Cat. No. CRL-1573), Jurkat (acute T-cell leukemia, Cat. No. TIB-152), K-562 (chronic myelogenous leukemia, Cat. No. CCL-243), HepG2 (liver carcinoma, Cat. No. HB-8065), T24 (urinary bladder carcinoma, Cat. No. HTB-4), SK-N-DZ (neuroblastoma, Cat. No. CRL-2149), NK-92 (non-Hodgkin's lymphoma, Cat. No. CRL-2407), MG-63 (osteosarcoma, Cat. No. CRL-1427), DU 145 (prostatic carcinoma, Cat. No. HTB-81), G-361 (melanoma, Cat. No. CRL-1424), PANC-1 (pancreatic carcinoma, Cat. No. CRL-1469), ES-2 (ovary

clear cell carcinoma, Cat. No. CRL-1978), Y79 (retinoblastoma, Cat. No. HTB-18), HT-29 (colorectal adenocarcinoma, Cat. No. HTB-38), H1299 (large cell lung carcinoma, Cat. No. CRL-5803), SNU1 (gastric carcinoma, Cat. No. CRL-5971), NL564 (EBV-transformed human lymphoblasts) and MCF10 (benign tumor breast cells).

RNA purification - Total RNA was extracted from the above mentioned human cell lines using TriReagent (Molecular Research Center, Cincinnati, OH). Poly(A)+ mRNA was purified using two cycles of the Dynabeads mRNA Purification Kit (Dynal Biotech ASA, Oslo, Norway), as per manufacturer instructions. The removal of traces of ribosomal RNA was confirmed by agarose gel electrophoresis. Poly(A)+ mRNAs from human testis, placenta, lung and brain tissue were purchased from BioChain Institute, Inc. (Hayward, CA). mRNAs of all cell lines described above were combined in equal quantities to obtain the reference 'mRNA pool'.

Preparation of labeled cDNA - For each hybridization, labeled cDNA was synthesized by reverse transcription of 0.5 µg of mRNA, in the presence of 100 pmol of random 9-mers, 1µg of oligo(dT)20, 1X RT buffer, 10 mM DTT, 3 nmol of Cy5- or Cy3-conjugated dUTP, 0.5 mM of dATP, dGTP and dCTP, and 0.2 mM dTTP, in a final volume of 40 µl (Amersham). The reaction mixture was incubated for 5 minutes at 65 °C and cooled to 42 °C. 600 Units of reverse transcriptase (Superscript II, Invitrogen, Carlsbad, CA) and 40 U of Rnase inhibitor (RNasin Promega, Madison, WI) were added and the reaction was incubated for 30 minutes at 42 °C. An additional 200 U of Superscript II were added and the reaction was incubated for another 15 minutes. Remaining RNA was degraded by the addition of 200 mM NaOH and 50 mM EDTA, at 65 °C for 10 minutes. The mixture was neutralized by adding half a volume of 1M Tris-HCl pH 7.5. Hybridizations were performed in duplicate using fluorescent reversal of Cy3- and Cy5-labeled cDNA from test cell mRNAs and pooled mRNAs. Pairs of Cy5/Cy3-labeled cDNA samples were combined, and

subsequently purified and concentrated to a final volume of 5-7 µl using a Microcon-30 (Millipore) concentrator.

Hybridization and washing conditions - Microarray slides were prehybridized with 40 µl of 5X SSC, 0.1 % SDS and 1 % BSA for 30 min at 42 °C, washed for 2 minutes with double distilled water, then rinsed with isopropanol, and spun dried at 500 g for 3 minutes. Prior to hybridization, the labeled probe was combined with 10 µg of Cot-1 DNA, 10 µg poly(dA)80, and 4 µg yeast tRNA, in a final volume of 15 µl. The mixture was denatured at 100 °C for 3 minutes and placed on ice. Formamide (final concentration 16 %), SSC (to 5X concentration) and 0.1 % SDS were added to a final volume of 30 µl. The mixture was placed on the array under a glass cover slip in a tightly sealed hybridization chamber, and immersed in a water bath at 42 °C, for 16 hours. Microarray slides were then washed for 4 minutes with 2X SSC, 0.1 % SDS; 4 minutes with 1X SSC, 0.01 % SDS; 4 minutes with 0.2X SSC and 15 seconds with 0.05X SSC and spun dry by centrifugation for 3 minutes at 500g.

Image processing - Following hybridization, arrays were scanned using a GenePix 4000B scanner (Axon Instruments, Union City, CA). Scanned array images were manually inspected and areas with visible artifacts or deformities were marked. Images were processed using GenePix Pro 3.0 (www.axon.com) software.

Normalization - The intensity for each spot was calculated as its mean intensity minus the median background around the spot. The signal for each oligo was calculated as the average of intensity values of the four redundant spots of each oligo. Normalization of the oligo signals was performed at several levels as is further described below.

Normalization of blocks was carried out in order to normalize the gradient of intensities within each slide. For each block i, an Ai parameter was calculated as the average of intensities of 56 positive control spots (oligonucleotide probes for the ubiquitously expressed housekeeping genes gapdh, actin, hsp70 and gnb2l1, in various probe concentrations). An average

A of all Ai averages was calculated. Based on this, a block normalization factor Bi was calculated for each block, as Bi = A/Ai, and applied to each spot in the block.

Normalization between slides was performed to bring all experiments to the same scale. For each experiment, the average of intensities of the 192 negative control spots on the array was set to be the 0 (zero) of the new scale. For a subset of highly signaling oligos, with intensities between the 70th and the 95th percentiles of the oligo signal distribution of the experiment, the average was arbitrarily set to be 500 in the new scale. The intensity of each oligo signal was accordingly converted to this new scale, to obtain the normalized signal. A ratio between the normalized cell-line signal and the normalized pool signal was calculated for each oligo in each experiment. To avoid misleading ratios coming from signals that were too low, the ratio Rji for oligo j in experiment i was calculated as: Rji = max [100, cell-line-signalji]/max [100, pool-signalji].

To normalize between red/green intensities in reciprocal experiments, the ratio Rjk for oligo j in cell-line k was calculated as the average of calculated ratios Rji between the two reciprocal experiments of the cell-line k. In cases where only one of the two reciprocal experiments showed an elevated or decreased ratio, while in the other the ratio was 1.0, the average Rjk was converted to 1.0.

The actual pool signal for each oligo was calculated to be the average of the normalized oligo signals in the pool channel of all experiments. A virtual pool signal was calculated as the average of the normalized oligo signals in the cell-line channel of all experiments. The virtual pool signals were found to be very close to the actual pool signals, indicating consistency in the analysis.

Threshold determination - To determine an expression threshold above, in which a normalized signal would be considered a 'positive' signal indicating expression, the distribution of all 16,512 normalized negative control signals and the standard deviation (neg-std-dev) were calculated. The neg-std-dev

obtained was 38. An oligo j was considered 'present' in a cell-line k if Rjk x actual-pool-signalj ≥ 4 x neg-std-dev.

EXAMPLE 1

Identification of 53BP1 and 76P RNA transcripts in a variety of human tissues and cell-lines

Background:

The tumor suppressor p53 binding protein 1 (SEQ ID NO: 15) is one of the various p53 target proteins. It binds to the DNA-binding domain of p53 and enhances p53-mediated transcriptional activation. 53BP1 is characterized by several structural motifs shared by several proteins involved in DNA repair and/or DNA damage-signaling pathways. 53BP1 becomes hyperphosphorylated and forms discrete nuclear foci in response to DNA damage induced by radiation and chemotherapy. Recent reports suggest that 53BP1 is an ataxia telangiectasia mutated (ATM) substrate that is involved early in the DNA damage-signaling pathways in mammalian cells, attributing a role to 53BP1 in the development of various mammalian pathologies.

Results:

Two 53BP1 RNA sense transcripts with dissimilar 3' UTRs were previously described [Iwabuchi K. et al. (1994) Proc. Natl. Acad. Sci. USA] and are illustrated in Figure 6 (red and green). Leads™ assembly program modified to uncover novel antisense transcripts was used to uncover three such transcripts for the 53BP1 gene, which transcripts have different 3' UTRs (SEQ ID NO: 16, 37 and 38) and encode the 76p gene product (Genbank accession number NM014444)(illustrated in blue).

To confirm expression of computationally retrieved antisense transcripts, two RNA-probes were generated. Schematic location of the probes used for sense and antisense validation (Riboprobe#1 and Riboprobe#2, respectively SEQ ID NO: 17 and 18, respectively) is illustrated in Figure 6. These RNA probes were used to identify the corresponding full-length transcripts.

As shown in Figure 7, Riboprobe#1 detected two transcripts of approximately 6.3 Kb and 10.5 Kb, corresponding to the sense mRNA. The absolute levels of the short messenger were rather homogeneous in all cell-lines examined. The 10.5 Kb variant exhibited a more heterogenic pattern of cellular distribution, and was mostly expressed in K562, MG-63, 293 HEK and Hela cells. In general, the longer sense transcript which is an alternatively polyadenylated variant was markedly lower expressed in the various cell lines examined.

The same membrane was used to perform northern analysis with Riboprobe#2 in order to validate expression of antisense transcripts of 53BP1. Results are shown in Figure 8. Three variants corresponding to the 76p gene were detected in most of the cell lines: 6.8 Kb, 4.2 Kb and 2.5 Kb. Minor fluctuations of expression were observed and the largest transcript was expressed at significantly higher levels than the smaller transcripts.

A sense strand probe was used to detect expression of the antisense transcripts in a variety of human tissues (Figure 9). The three alternatively polyadenylated variants with different 3' UTRs were expressed in most of the tissues. Total levels of these transcripts varied in the different tissues assayed. For example, highest level of expression for all three transcripts was observed in the brain and testis, while no expression of the 6.8 Kb and 4.2 Kb variants was detected in the spleen. Expression levels of each transcript were summarized in Table 2 below.

Table 2

	Transcript Mol. Weight (Kb)		
Tissue	6.8	4.2	2.5
brain	+++	++++	++++
colon	+	++	+
heart	-	+	++
kidney	++	++	+
Liver	-	-	+
lung	++++	+++	+

muscle	++	+	+
placenta	+	++	++
Small intestine.	++	++	-
spleen	-	-	+
stomach	-	-	+
testis	++	++	++++

Reverse transcription amplification (RT-PCR) analysis was performed in order to substantiate the northern blot results. Primers were synthesized according to the scheme shown in Figure 10 (indicated by arrows). The expected amplification products corresponded completely to the observed amplification reaction products, supporting the existence of the various 53BP1 and 76p transcription variants.

EXAMPLE 2

Identification of mRNA and complementary transcripts of the Cell death inducing DFF45-like effector (CIDE)-B

Background:

Cell death inducing DFF45-like effector (CIDE-B) (GenBank Accession numbers AF190901 and AF218586) is a member of a novel family of apoptosis-inducing factors that share homology with the N-terminal region of DFF, the DNA fragmentation factor. Although the molecular mechanism of CIDE-B induced apoptosis in unclear, mitochondrial localization and dimerization, both where shown to be required [Chen Z. et al. (2000) J. Biol. Chem. 275:22619-22622]. Notably, over-expression of CIDE-B in mammalian cells shows strong cell death-inducing activity, suggesting that aberrant expression of this protein may be associated with a number of mammalian pathologies [Inohara N. et al. (1998) EMBO J. 17:2526-2533].

Results:

Two sense transcript of the CIDE-B gene were previously described with different 5' UTRs [Inohara N. et al. (1998) EMBO J. 17:2526-2533 and

Lugovskoy AA. et al. (1999) Cell 99:745-755] (SEQ ID NOs: 19 and 20). Computational analysis recovered a potential elongated BLTR2 transcript (SEQ ID NO: 21), showing full complementary to the CIDE-B mRNA transcripts (Figure 11).

Northern blot analysis was done in order to determine the distribution of the CIDE-B sense and antisense transcripts in various cell-lines. A 430 base pairs DNA fragment was selected to generate RNA probes for identification of both sense and antisense transcripts (SEQ ID NOs: 22 and 23, respectively).

Expression of antisense mRNA transcripts was detected in various cell-lines and especially in the mammary gland adenocarcinome cell line-MCF-7 as a predominant 6.5 Kb transcript, although higher forms were also visualized (Figure 12). Low hybridization with a CIDE-B probe was detected (Figure 13).

Conclusion:

BLTR2 was recently identified as a putative seven-transmembrane receptor with a high homology to the Leukotriene B (4) receptor [Tryselius Y. et al. (2000) Biochem. Biophys. Res. Commun. 274:377-82]. Although the mechanism of action of BLTR2 is poorly understood, it is conceivable that BLTR2 mRNA plays a role in the regulation of CIDE-B apoptotic effector and vice versa.

EXAMPLE 3

Identification of mRNA and complementary transcripts of the apoptosis inducing factor APAF-1

Background:

A conserved series of events including cellular shrinkage, nuclear condensation, externalization of plasma membrane phosphatidyl serine, and oligonucleosomal DNA fragmentation characterizes apoptotic cell death. Regardless of the circumstance, induction and execution of apoptotic events require activation of caspases, a family of aspartate-specific cysteine proteinases. Caspase activation may be regulated by the mitochondrion and

specifically by the apoptosome consisting of an oligomeric complex of apoptotic protease-activating factor-1 (APAF-1), cytochrome C and dATP. The apoptosome recruits and activates caspase-9, which in turn activates the executioner caspases, caspase-3 and -7. The active executioners kill the cell by proteolysis of key cellular substrates [Zou H. et al. (1999) J. Biol. Chem. 274:11549-11556]. Evasion or inactivation of the mitochondrial apoptosis pathway may contribute to oncogenesis by allowing cell proliferation. In this instance, unregulated cell proliferation may occur by inactivation of APAF-1, which has been suggested to occur via genetic loss or inhibition by HSP-70 and HSP-90. Although aberrant expression of APAF-1 was found in a variety of malignancies (including ovarian epithelial cancer), no link was found to accelerated protein degradation.

Results:

One RNA transcript has been previously described for APAF-1 [Zou H. et al. (1999) J. Biol. Chem. 274:11549-11556] (SEQ ID NO: 10) (SEQ ID NO: 24). Computational search for natural antisense transcripts has revealed two complementary transcripts for APAF-1 messenger RNA (SEQ ID NOs: 25 and 26). These antisense transcripts include an open reading frame encoding the EB-1 gene (GenBank accession numbers AF145204; AF164792). The overlap between the APAF-1 messenger RNA and the longer antisense transcript is of at least 300 nucleotides.

To validate expression of computationally retrieved antisense transcripts for APAF-1, as well as expression of APAF-1 mRNA in the assayed human cell lines, RNA-probes of 366 ribonucleotides were generated (sense and antisense strands, respectively). Schematic location of the probes used for sense and antisense validation (Riboprobe#1 and Riboprobe#2, SEQ ID NOs: 27 and 28, respectively) is illustrated in Figure 14.

As shown in Figure 15a, the sense RNA probe directed at visualizing the antisense transcripts, identified a clear band of 3 Kb corresponding to the long computationally retrieved antisense transcript as well as other transcripts sizing

from 1 Kb to 8 Kb (Figure 15a). Transcripts were essentially found in all cell lines but especially in 293 HEK and LN-Cap lines.

Hybridization with an RNA probe directed at visualizing the mRNA transcript of APAF-1 resulted only in a blurred patterns (Figure 15b). However, a 7 Kb mRNA transcript consistent with APAF-1mRNA was seen in Ln Cap and 293 HEK cell lines.

Conclusion:

A reciprocal pattern of expression was observed for both APAF-1 and EB-1 transcripts, exhibiting an interesting expressional relationship between the sense and antisense transcripts suggesting antisense-mediated expression regulation.

EXAMPLE 4

mRNA expression of muscle nicotinic Acetyl-Choline Receptor ε subunit and its complementary MINK transcript

Background:

The muscle nicotinic Acetylcholine Receptor ε subunit (AChRε) encodes for one of five subunits of a ligand gated ion channel receptor located at the neuromuscular synapse. AChRε is up-regulated in the postnatal period when it replaces γ subunit of the receptor [Witzamann, V. et al., (1987) FEBS Lett. 223, 104-112]. It is also up-regulated in synapse development, specifically by the trophic factor neuregulin [Martinou J. C. (1991) Pro. Natl. Acad. Sci. USA 88, 7669-7673]. In an attempt to decipher AchRε function and mechanism of regulation, computational screen for AChRε K complementary transcript was carried out.

Results:

One mRNA transcript of AChRε gene was previously described [Beeson D. Eur. J. Biochem (1993) 215, 229-238] (SEQ ID NO: 29). Computational analysis recovered a complementary transcript belonging to Mink, a new member of the germinal center kinase (GCK) family (SEQ ID NO: 30) [Dan I.

FEBS Lett. (2000) 469, 19-23] showing an overlap of at least 280 nucleotides to the AchRε mRNA, as schematically illustrated in Figure 16.

To validate the overlap of the two genes and to learn about their tissue distribution, northern analysis of a variety of human tissues was performed. Poly(A)-RNA containing membrane was hybridized with a 280 nucleotides RNA probes, corresponding to the overlap region in either antisense or sense orientation (SEQ ID NOs: 31 and 32, respectively).

As is evident from Figure 17a an AChRε transcript was expressed as a predominant 4 Kb band and had the highest expression in the heart, kidney and brain while surprisingly only a limited expression was observed in the skeletal muscle.

Hybridization with a MINK specific RNA probe revealed a major transcript of about 5 Kb, in accordance with previous results [Dan I. FEBS Lett. (2000) 469, 19-23] (Figure 17b). The mRNA transcript was ubiquitously expressed with strongest expression found in brain, liver, thymus, spleen and pancreas, again in agreement with Dan I. et al.

Conclusion:

The finding that AChRε and Mink genes are antisense each to one another with a significant overlap, and the fact that the two genes are co-expressed in some tissues (eg., brain) suggest the possibility that one of them may regulate the other under certain conditions.

EXAMPLE 5

Expression of Cyclin E2 mRNA and complementary transcripts in a variety of human cell-lines

Background:

The human cyclin E2 gene encodes a 404-amino-acid protein that is most closely related to cyclin E. Cyclin E2 associates with Cdk2 in a functional kinase complex that is inhibited by both p27(Kip1) and p21(Cip1). The catalytic activity associated with cyclin E2 complexes is cell cycle regulated

and peaks at the G1/S transition. Overexpression of cyclin E2 in mammalian cells accelerates cell-cycle progression. Unlike cyclin E1, cyclin E2 levels are low to undetectable in nontransformed cells and increase significantly in tumor-derived cells suggesting specific mechanism of regulation.

Results:

One RNA transcript was found for cyclin E2 (SEQ ID NO: 33. Computational search for natural antisense transcripts has revealed one complementary transcript for cyclin E2 messenger RNA (SEQ ID NO: 34). The overlap between the cyclin E2 sense RNA and the antisense transcript is of at least 72 nucleotides.

To confirm expression of the computationally retrieved antisense transcript for cyclin E2 as well as of cyclin E2 mRNA in human cell lines, two RNA-probes of 800 ribonucleotides were generated. Schematic location of the probes used for sense and antisense validation (SEQ ID NO: 44, Riboprobe#1 is illustrated in Figure 18).

As shown in Figure 19a, Riboprobe#1 detected two transcripts of approximately 3 Kb and 4.3 Kb. The absolute levels of the transcripts were quite heterogenic in all cell-lines examined. Both transcripts were completely absent from the Ln Cap cell line, while significantly high expression was observed in MCF-7 and DLD-1 lines, especially of the short transcript.

The same membrane was used to perform northern analysis with Riboprobe#2 in order to validate expression of antisense transcripts of cyclin E2. As is evident from Figure 19b, an antisense transcript 3.8 Kb long was observed in most cells assayed. Significantly high pattern of expression was observed in K562, MCF-7 and DLD-1 cell lines, while only a very moderate level of expression was detected in Ln Cap and HepG2 cell lines.

EXAMPLE 6

Co-regulated expression of CIDE-B and its complementary transcript upon induction of apoptosis

The discovery of a novel naturally occurring antisense transcript to the apoptosis inducing factor, CIDE-B (see Example 2 hereinabove), suggested that the latter may be regulated by its complementary transcript, thereby establishing a novel mechanism of regulation. To address this, differential expression analysis of CIDE-B expression and its endogenous antisense transcript expression was performed following induction of apoptosis.

Materials and methods

Induction of apoptosis and reverse transcription analysis -

Monolayers of 293 cells were either left untreated (UT) or incubated with increasing concentrations of etoposide or staurosporine (Sigma IL). Twenty-four hours following addition of the drug, total RNA was extracted as decribed hereinabove. Purified RNA was further treated with DNaseI. A reverse transcription reaction were carried out with equivalent amounts of RNA in a final volume of 20 μl containing 100 pmol of the oligo(dT) primer, 250 ng of total RNA, 0.5 mM each of four deoxynucleoside triphosphates and 5 units of reverse transcriptase. The reaction mixture was incubated at 65 °C for 5 min, 42 °C for 50 min and 70 °C for 15 min. PCR was carried out in a final volume of 25 μl containing 12.5 pmol each of the oligonucleotide primers derived of exons 3 and 7 of CIDE-B (SEQ ID NOs: 39 and 40), 1 μl of RT solution and 1.75 units of *Taq* polymerase. Amplification was carried out by an initial denaturation step at 94 °C for 5 min followed by 35 cycles of [94 °C for 30 s, 68 °C for 30 s, and 68 °C for 130 min]. At the end of the PCR amplification, products were analyzed on agarose gels stained with ethidium bromide and visualized with UV light.

Results

Amplification reaction yielded two major PCR products of 740 bp and 2285 bp (Figure 20). The small (740 bp) PCR product derived from the sense

(CIDE-B) strand, whereas the larger (2285 bp) product represented an intronless antisense transcript. Evidently, an increase of sense transcript, concomitant with a decrease of antisense transcript, was observed following treatment with etoposide (lanes 1-4) as compared to untreated cells (lane 9), while no change was detected following staurosporine treatment (lanes 5-8).

These results suggest that following induction of apoptosis, antisense regulation of CIDE-B is abolished thereby allowing CIDE-B mediated apoptosis to proceed.

EXAMPLE 7

Reciprocal variation in sense and antisense expression of mouse nicotinic acetylcholine receptor, epsilon subunit during differentiation

The mouse nicotinic acetylcholine receptor, epsilon (mAchRε) subunit (SEQ ID NO: 35) has a critical function in a variety of differentiation processes. To address a novel concept of antisense regulation of AchRε-mediated differentiation, expression patterns of AchRε and its naturally occurring antisense transcript (SEQ ID NO: 36) were examined following induction of differentiation.

Materials and methods

Induction of apoptosis and reverse transcription analysis - C2 mouse myoblast cells were incubated with a differentiation medium (Dulbecco's modified Eagle's medium (DMEM) including 10 μg/ml insulin and 10 μg/ml transferring) or control medium (untreated) for 48 and 72 hours. Total RNA was extracted from treated and control cells and reverse-transcribed. PCR was done using F4 and R3 primers, derived from exon numbers 10 and 12 (last exon, SEQ ID NOs: 41 and 42, respectively) of the mouse nicotinic acetylcholine receptor, epsilon subunit (mAChRε) and directed at detecting sense and antisense transcripts (see Figure 21a).

Results

Amplification reaction showed a gradual increase in AchRε transcript expression, concomitant with the differentiation state of the cells. A second amplification product, which corresponded to an unspliced transcript was seen in untreated cells and disappeared following induction of differentiation. This fragment corresponds to a putative antisense transcript of the AchRε, and may represent an alternative 3' UTR of the Mink gene , of which the known transcript terminates 400 bp downstream to AchRε (see Example 4). To overcome possible competition between the two transcripts, another PCR reaction was carried out using antisense specific riboprobes F4 and R4 (SEQ ID NO: 43). Reverse transcription products of this amlification reaction showed a single band which corresponded to a naturally occurring antisense transcript of the AchRε. As expected this transcript disappeared following induction of differentiation.

These results imply inverse regulation of the AchRε and its naturally occurring antisense transcript, during muscle cells differentiation from myoblasts to myotubes. Regulation may proceed, possibly through complementation of the sense and antisense transcripts to form dsRNA which can serve as a substrate for double strand RNA processing enzymes such as RNase H.

EXAMPLE 8

A polynucleotide database of sequences corresponding to the naturally occurring antisense transcripts identified by the present invention and their complementary sense sequences

Naturally occurring antisense sequences identified according to the teachings of the present invention and their corresponding sense sequences are provided in the CD-ROMs enclosed herewith (file content: CD-ROM1 includes a "seq" text file which contains the actual polynucleotide sequences, and a "table" file which contains summarized data pertaining to each sense-

antisense sequence pair. CD-ROM2 includes an "aligments" file which contains sequence alignments of sense and antisense overlapping regions. CD-ROM3 contains Excel files: "Table S1" and "Table S2", further described in Example 9.

Table 3 below exemplifies the format of the Table provided in CD-ROM1. Each row represents a pair of transcripts. The columns of Table 3 represent (from the left): the serial number of the pair, the name of the first transcript, its length in nucleotides, the name of the second transcript, its length in nucleotides, the number of base pairs that overlap between the two transcripts, offsets of overlap beginning at the first transcript, offsets of overlap beginning at the second transcript.

Table 3

Serial No.	First transcript	First transcript length (nt)	Second transcript	Second transcript length (nt)	Overlap length (nt)	Start of overlap in first / in second transcript
570_0	AV705532_0 (SEQ ID NO: 1)	190	Z44352_15 (SEQ ID NO: 2)	783	OL: 52	OF1: 1 OF2: 1
570_1	AV705532_0	190	Z44352_14 (SEQ ID NO: 3)	1649	OL: 52	OF1: 1 OF2: 1
570_2	AV705532_0	190	Z44352_13 (SEQ ID NO: 4)	1861	OL: 52	OF1: 1 OF2: 1
571_0	AW070860_0 (SEQ ID NO: 5)	214	T81142_7 (SEQ ID NO: 6)	1934	OL: 54	OF1: 1 OF2: 1162
571_1	AW070860_0	214	T81142_6 (SEQ ID NO: 7)	2353	OL: 54	OF1: 1 OF2: 1162
571_2	AW070860_0	214	T81142_4 (SEQ ID NO: 8)	2500	OL: 54	OF1: 1 OF2: 1264
571_3	AW070860_0	214	T81142_3 (SEQ ID NO: 9)	947	OL: 54	OF1: 1 OF2: 171
571_4	AW070860_0	214	T81142_2 (SEQ ID NO: 10)	1366	OL: 54	OF1: 1 OF2: 171
572_0	BE046369_0 (SEQ ID NO: 11)	422	W26553_3 (SEQ ID NO: 12)	1532	OL: 52	OF1: 1 OF2: 1532
572_1	BE046369_0	422	W26553_2 (SEQ ID NO: 13)	1753	OL: 52	OF1: 1 OF2: 1753

572_2	BE046369_0	422	W26553_1	1832	OL: 52	OF1: 1 OF2: 1832
			(SEQ ID NO: 14)			

Pairs of transcripts are numbered, (within a contig pair, right to the underscore) that belong to a pair of contigs (numbered left to the underscore). Transcript names are arbitrary designataions.

Sequence alignment of the overlapping region in each sense and antisense pair of Table 1 is demonstrated in Figure 4a-k. Alignments were performed using the BLAST sequence alignment algorithm (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST). Interestingly, alignment profile shows high level of variability with regard to overlap lengths. It is conceivable that short overlaps are due to technical reasons associated with insufficient sequence data.

The putative naturally occurring antisense transcripts identified by the present invention and disclosed in the enclosed CD-ROMs can be used to detect and/or treat a variety of diseases, disorders or conditions, examples of which are listed hereinunder. For example, antisense transcripts or sequence information derived therefrom can be used to construct microarray kits (described in details in the preferred embodiments section) dedicated to diagnosing specific diseases, disorders or conditions.

The following sections list examples of proteins (subsection i), based on their molecular function, which participate in variety of diseases (listed in subsection ii), which diseases can be diagnosed/treated using information derived from naturally occurring antisense transcripts such as those uncovered by the present invention.

i. Molecular function

defense/immunity proteins

Information derived from proteins involved in the immune and complement systems, such as acute-phase response proteins, antimicrobial peptides, antiviral response proteins, blood coagulation factors, complement components, immunoglobulins, major histocompatibility complex antigens, and opsonins can be used to diagnose/treat diseases involving the immunological

system including inflammation, autoimmune diseases, infectious diseases, as well as cancerous processes. Diseases which are manifested by non-normal coagulation processes, which may include abnormal bleeding or excessive coagulation.

Immunoglobulins

Information derived from proteins involved in the immune and complement systems including antigens and autoantigens, immunoglobulins, MHC and HLA proteins and their associated proteins can be used to diagnose/treat diseases involving the immunological system including inflammation, autoimmune diseases, infectious diseases, as well as cancerous processes.

Nucleotide binding proteins

Information derived from ligand binding or carrier proteins can be used to diagnose/treat diseases involving dysregulated expression, activity or localization of nucleotide binding proteins.

Nucleic acid binding proteins

Information derived from proteins involved in RNA and DNA synthesis and expression regulation, such as transcription factors, RNA and DNA binding proteins, zinc fingers, helicase, isomerase, histones, nucleases, ribonucleoproteins, transcription and translation factors and others can be used to diagnose/treat diseases involving DNA or RNA binding proteins such as: helicases, isomerases, histones and nucleases, for example diseases where there is non-normal replication or transcription of DNA and RNA respectively.

RNA polymerase II transcription factors

Information derived from proteins such as specific and non-specific RNA polymerase II transcription factors, enhancer binding, ligand-regulated transcription factor and general RNA polymerase II transcription factors can be used to diagnose/treat diseases involving RNA polymerase II transcription factors, for example disorders involving abnormal transcription of RNA.

RNA binding proteins

Information derived from RNA binding proteins involved in splicing and translation regulation, such as tRNA binding proteins, RNA helicases, double-stranded RNA and single-stranded RNA binding proteins, mRNA binding proteins, snRNA cap binding proteins, 5S RNA and 7S RNA binding proteins, poly-pyrimidine tract binding proteins, snRNA binding proteins, and AU-specific RNA binding proteins can be used to diagnose/treat diseases involving transcription and translation factors such as: helicases, isomerases, histones and nucleases, for example diseases where there is non-normal transcription, splicing, post-transcriptional processing, translation or stability of the RNA.

Chaperones

Information derived from proteins such as ribosomal chaperone, peptidylprolyl isomerase, lectin-binding chaperone, nucleosome assembly chaperone, chaperonin ATPase, cochaperone, heat shock protein, HSP70/HSP90 organizing protein, fimbrial chaperone, metallochaperone, tubulin folding, HSC70-interacting protein can be used to diagnose/treat diseases involving pathological conditions, which are associated with non-normal protein activity or structure. Binding of the products of the proteins of this family, or antibodies reactive therewith, can modulate a plurality of protein activities as well as change protein structure. Alternatively, diseases in which there is abnormal degradation of other proteins, which may cause non-normal accumulation of various proteinaceous products in cells, caused non- normal (prolonged or shortened) activity of proteins, etc.

Motor proteins

Information derived from proteins that generate force or energy by the hydrolysis of ATP and that function in the production of intracellular movement or transportation including microfilament motor, axonemal motor, microtubule motor, kinetochore motor (like dynein, kinesin, or myosin) can be used to diagnose/treat diseases involving un-normal chemotactic movement or

motor dependent macromolecule operation such as of dynamin, which affects the regulated endocytic process.

Actin binding proteins

Information derived from actin binding proteins, such as actin cross-linking, actin bundling, F-actin capping, actin monomer binding, actin lateral binding, actin depolymerizing, actin monomer sequestering, actin filament severing, actin modulating, membrane associated actin binding, actin thin filament length regulation and actin polymerizing proteins can be used to diagnose/treat diseases involving cytoskeletal malformations, aberrant cellular morphology affecting extracellular interactions and dysregulated intracellular signaling.

Enzymes

Information derived from proteins possessing enzymatic activities, such as mannosylphosphate transferase, para-hydroxybenzoate:polyprenyltransferase, Rieske iron-sulfur protein, imidazoleglycerol-phosphate synthase, sphingosine hydroxylase , tRNA 2'-phosphotransferase, sterol C-24(28) reductase, C-8 sterol isomerase, C-22 sterol desaturase, C-14 sterol reductase, C-3 sterol dehydrogenase (C-4 sterol decarboxylase) , 3-keto sterol reductase, C-4 methyl sterol oxidase, dihydronicotinamide riboside quinone reductase, glutamate phosphate reductase, DNA repair enzyme, telomerase, alpha-ketoacid dehydrogenase, beta-alanyl-dopamine synthase, RNA editase, aldo-keto reductase, alkylbase DNA glycosidase, glycogen debranching enzyme, dihydropterin deaminase, dihydropterin oxidase, dimethylnitrosamine demethylase, ecdysteroid UDP-glucosyl/UDP glucuronosyl transferase, glycine cleavage system, helicase, histone deacetylase, mevaldate reductase, monooxygenase, poly(ADP-ribose) glycohydrolase, pyruvate dehydrogenase, serine esterase, sterol carrier protein X-related thiolase, transposase , tyramine-beta hydroxylase, para-aminobenzoic acid (PABA) synthase, glu-tRNA(gln) amidotransferase, molybdopterin cofactor sulfurase, lanosterol 14-alpha-demethylase, aromatase,

4-hydroxybenzoate octaprenyltransferase, 7,8-dihydro-8-oxoguanine-triphosphatase, CDP-alcohol phosphotransferase, 2,5-diamino-6-(ribosylamino)-4(3H)-pyrimidonone 5'-phosphate deaminase, diphosphoinositol polyphosphate phosphohydrolase, gamma-glutamyl carboxylase, small protein conjugating enzyme, small protein activating enzyme, 1-deoxyxylulose-5-phosphate synthase, 2'-phosphotransferase, 2-octoprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, 3,4 dihydroxy-2-butanone-4-phosphate synthase, 4-amino-4-deoxychorismate lyase, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase, ADP-L-glycero-D-manno-heptose synthase, D-erythro-7,8-dihydroneopterin triphosphate 2'-epimerase, N-ethylmaleimide reductase, O-antigen ligase, O-antigen polymerase, UDP-2,3-diacylglucosamine hydrolase, arsenate reductase, carnitine racemase, cobalamin [5'-phosphate] synthase , cobinamide phosphate guanylyltransferase, enterobactin synthetase, enterochelin esterase, enterochelin synthetase, glycolate oxidase, integrase, lauroyl transferase, peptidoglycan synthetase, phosphopantetheinyltransferase, phosphoglucosamine mutase, phosphoheptose isomerase, quinolinate synthase, siroheme synthase, N-acylmannosamine-6-phosphate 2-epimerase, N-acetyl-anhydromuramoyl-L-alanine amidase, carbon-phosphorous lyase, heme-copper terminal oxidase, disulfide oxidoreductase, phthalate dioxygenase reductase, sphingosine-1-phosphate lyase, molybdopterin oxidoreductase, dehydrogenase, NADPH oxidase, naringenin-chalcone synthase, N-ethylammeline chlorohydrolase, polyketide synthase, aldolase, kinase, phosphatase, CoA-ligase, oxidoreductase, transferase, hydrolase, lyase isomerase, ligase, ATPase, sulfhydryl oxidase, lipoate-protein ligase, delta-1-pyrroline-5-carboxyate synthetase, lipoic acid synthase and tRNA dihydrouridine synthase can be used to diagnose/treat diseases which can be ameliorated by modulating the activity of various enzymes which are involved both in enzymatic processes inside cells as well as in cell signaling.

Protein serine/threonine kinases

Information derived from kinases, which phosphorilate serine/threonine residues, mainly involved in signal transduction, such as transmembrane receptor protein serine/threonine kinase, 3-phosphoinositide-dependent protein kinase, DNA-dependent protein kinase, G-protein-coupled receptor phosphorylating protein kinase, SNF1A/AMP-activated protein kinase, casein kinase, calmodulin regulated protein kinase, cyclic-nucleotide dependent protein kinase, cyclin-dependent protein kinase, eukaryotic translation initiation factor 2alpha kinase, galactosyltransferase-associated kinase, glycogen synthase kinase 3, protein kinase C, receptor signaling protein serine/threonine kinase, ribosomal protein S6 kinase and IkB kinase can be used to treat, or detect, respectively, diseases which may be ameliorated by a modulating kinase activity, which is one of the main signaling pathways inside cell.

Enzyme inhibitors

Information derived from inhibitors and suppressors of other proteins and enzymes, such as inhibitors of Kinases, phosphatases, chaperones, guanylate cyclase, DNA gyrase, ribonuclease, proteasome inhibitors, diazepam-binding inhibitor, ornithine decarboxylase inhibitor GTPase inhibitors, dUTP pyrophosphatase inhibitor, phospholipase inhibitor, proteinase inhibitor, protein biosynthesis inhibitors, alpha-amylase inhibitors can be used to treat diseases in which beneficial effect may be achieved by modulating the activity of inhibitors and suppressors of proteins and enzymes.

Signal transducers

Information derived from various signal transducers, such as activin inhibitors, receptor-associated proteins alpha-2 macroglobulin receptors, morphogens, quorum sensing signal generators, quorum sensing response regulators, receptor signaling proteins, ligands, receptors, two-component sensor molecules, two-component response regulators can be used to diagnose/treat diseases involving abnormal signal-transduction, either as a cause, or as a result of the disease.

Receptors

Information derived from various receptors, such as signal transducers, complement receptors, ligand-dependent nuclear receptors, transmembrane receptors, GPI-anchored membrane-bound receptors, various coreceptors, internalization receptors, receptors to neurotransmitters, hormones and various other effectors and ligands can be used to diagnose/treat diseases involving various receptors, including receptors to neurotransmitters, hormones and various other effectors and ligands.

Receptor signaling proteins

Information derived from receptor proteins involved in signal transduction, such as receptor signaling protein serine/threonine kinase, receptor signaling protein tyrosine kinase, receptor signaling protein tyrosine phosphatase, aryl hydrocarbon receptor nuclear translocator, hematopoeitin/interferon-class (D200-domain) cytokine receptor signal transducer, transmembrane receptor protein tyrosine kinase signaling protein, transmembrane receptor protein serine/threonine kinase signaling protein, receptor signaling protein serine/threonine kinase signaling protein, receptor signaling protein serine/threonine phosphatase signaling protein, small GTPase regulatory/interacting protein, receptor signaling protein tyrosine kinase signaling protein, and receptor signaling protein serine/threonine phosphatase can be used to diagnose/treat diseases involving non-normal signal transduction, either as a cause, or as a result of the disease.

Small GTPase regulatory/interacting proteins

Information derived from small GTPase regulatory proteins, such as RAB escort protein, guanyl-nucleotide exchange factor, guanyl-nucleotide exchange factor adaptor, GDP-dissociation inhibitor, GTPase inhibitor, GTPase activator, guanyl-nucleotide releasing factor , GDP-dissociation stimulator, regulator of G-protein signaling, RAS interactor, RHO interactor, RAB interactor, and RAL interactor can be used to diagnose/treat diseases involving

signal-transduction, typically involving G-proteases is non-normal, either as a cause, or as a result of the disease.

Ligands

Information derived from ligands such as opioid peptides, baboon receptor ligand, branchless receptor ligand, breathless receptor ligand, ephrin, frizzled receptor ligand, frizzled-2 receptor ligand, heartless receptor ligand, Notch receptor ligand, patched receptor ligand, punt receptor ligand, Ror receptor ligand, saxophone receptor ligand, SE20 receptor ligand, sevenless receptor ligand, smooth receptor ligand, thickveins receptor ligand, Toll receptor ligand, Torso receptor ligand, death receptor ligand, scavenger receptor ligand, neuroligin, integrin ligand, hormones, pheromones, growth factors and sulfonylurea receptor ligand can be used to diagnose/treat:

(a) diseases involving non-normal secretion of proteins, which may be due to non-normal presence, absence or non-normal response to normal levels of secreted proteins including hormones, neurotransmitters, and various other proteins secreted by cells to the extracellular environment;

(b) diseases which are endocrine in essence (cause or are a result of hormones), or may be ameliorated by raising, or decreasing the level of hormones and proteins;

(c) diseases which may be ameliorated by modulating the concentration or activity or interaction binding, etc. of growth factors, cytokines, interleukins, interferon and lymphokines, typically diseases such as autoimmune diseases, inflammation related disease, Graft vs. Host diseases, diseases caused by infectious agents, cancer diseases, as well as disease originating from improper concentration of growth factors causing non-normal (either excessive or too little of) growth of various tissues themselves, or causing untimely death of a desired cell population; and

(d) diseases which are manifested by non-normal development, which may be non-normal development of the organism (genetic diseases

involving non-normal development of a fetus), non-normal development of a tissue (a tissue which is not properly developed) as well as cancer diseases.

Cell adhesion molecules

Information derived from proteins that serve as adhesion molecules between adjoining cells, such as membrane-associated protein with guanylate kinase activity, cell adhesion receptor, neuroligin, calcium-dependent cell adhesion molecule, selectin, calcium-independent cell adhesion molecule, extracellular matrix protein can be used to diagnose/treat diseases where adhesion between adjoining cells is involved, typically conditions in which the adhesion is non-normal. Typical examples of such conditions are cancer conditions in which non-normal adhesion may cause and enhance the process of metastasis. Other examples of such conditions include conditions of non-normal growth and development of various tissues in which modulation adhesion among adjoining cells can improve the condition.

Structural proteins

Information derived from proteins involved in cell structure, such as ribosomal proteins, cell wall proteins, cytoskeletal proteins, extracellular matrix proteins, extracellular matrix glycoproteins, amyloid proteins, plasma proteins, eye lens proteins, chorion proteins (sensu Insecta), cuticle proteins (sensu Insecta), puparial glue protein (sensu Diptera), bone proteins, yolk proteins, muscle proteins, vitelline membrane proteins (sensu Insecta), peritrophic membrane proteins (sensu Insecta), and nuclear pore proteins can be used to diagnose/treat diseases involving abnormalities in cytoskeleton, including cancerous cells, and diseased cells including those which do not propagate, grow or function normally. Diseases involving non-normal sub-cellular proteins such as non-normal ribozymal proteins.

Transporter proteins

Information derived from proteins such as amine/polyamine transporter, lipid transporter, neurotransmitter transporter, organic acid transporter, oxygen transporter, water transporter, carriers, intracellular transportes, protein

transporters, ion transporters, carbohydrate transporter, polyol transporter, amino acid transporters, vitamin/cofactor transporters, siderophore transporter, drug transporter, channel/pore class transporter, group translocator, auxiliary transport proteins, Permeases, murein transporter, organic alcohol transporter, nucleobase, nucleoside and nucleotide and nucleic acid transporters can be used to diagnose/treat diseases in which abnormal transport of molecules and macromolecules such as neurotransmitters, hormones, sugar etc. leads to various pathologies.

Intracellular transporters

Information derived from proteins that mediate the transport of molecules and macromoleules inside the cell, such as intracellular nucleoside transporter, vacuolar assembly proteins, vesicle transporters, vesicle fusion proteins, and type II protein secretors can be used to diagnose/treat diseases in which abnormal transport of molecules and macromolecules leads to various pathologies.

Ligand binding or carrier proteins

Information derived from various proteins, involved in diverse biological functions, such as pyridoxal phosphate binding, carbohydrate binding, magnesium binding, amino acid binding, cyclosporin A binding, nickel binding, chlorophyll binding, biotin binding, penicillin binding, selenium binding, tocopherol binding, lipid binding, drug binding, oxygen transporter, electron transporter, steroid binding, juvenile hormone binding, retinoid binding, heavy metal binding, calcium binding, protein binding, glycosaminoglycan binding, folate binding, odorant binding, lipopolysaccharide binding, and nucleotide binding can be used to diagnose/treat diseases involving improper intracellular or extracellular accumulation or removal of small molecules such as calcium ions, improper incorporation of metals and modified amino acids (i.e., seleno-cystein), dysregulated signaling effected by improper steroid titration etc.

Electron transporters

Information derived from ligand binding proteins or carrier proteins involved in electron transport, such as flavin-containing electron transporter, cytochromes, electron donors, electron acceptors, electron carriers and cytochrome-c oxidases can be used to diagnose/treat diseases involving dysregulated mitochondrial activity.

Calcium binding proteins

Information derived from calcium binding proteins, ligand binding proteins or carriers, such as diacylglycerol kinase, Calpain, calcium-dependent protein serine/threonine phosphatase, calcium sensing proteins and calcium storage proteins can be used to diagnose/treat diseases in which intracellular or extracellular calcium storage or release is improper.

Binding proteins

Information derived from various proteins exhibiting intermediate filament binding, LIM-domain binding, LLR-domain binding, clathrin binding, ARF binding, vinculin binding, KU70 binding, troponin C binding PDZ-domain binding, SH3-domain binding, fibroblast growth factor binding, membrane-associated protein with guanylate kinase activity interacting, Wnt-protein binding, DEAD/H-box RNA helicase binding, beta-amyloid binding, myosin binding, TATA-binding protein binding DNA topoisomerase I binding, polypeptide hormone binding, RHO binding, FH1-domain binding, syntaxin-1 binding, HSC70-interacting, transcription factor binding, metarhodopsin binding, tubulin binding, JUN kinase binding, RAN protein binding, protein signal sequence binding, importin alpha export receptor, poly-glutamine tract binding, protein carrier, beta-catenin binding, protein C-terminus binding, lipoprotein binding, cytoskeletal protein binding protein, nuclear localization sequence binding, protein phosphatase 1 binding, adenylate cyclase binding, eukaryotic initiation factor 4E binding, calmodulin binding, collagen binding, insulin-like growth factor binding, lamin binding, profilin binding, tropomyosin binding, actin binding, peroxisome targeting sequence binding, SNARE binding

and cyclin binding can be used to diagnose/treat diseases involving non-normal protein activity or structure. Binding of the products of the variants of this family, or antibodies reactive therewith, can modulate a plurality of protein activities as well as change protein structure.

Transcription factor binding proteins

Information derived from proteins involved in transcription factors binding, RNA and DNA binding, such as transcription factors, RNA and DNA binding proteins, zinc fingers, helicase, isomerase, histones, and nucleases can be used to diagnose/treat diseases involving transcription factors binding proteins, for example diseases where there is abnormal replication or transcription of DNA and RNA respectively.

Enzyme regulators

Information derived from enzyme regulators, such as activators of kinases, phosphatases, sphingolipids, chaperones, guanylate cyclase, tryptophan hydroxylase, proteases, phospholipases, caspases, proprotein convertase 2 activator, cyclin-dependent protein kinase 5 activator, superoxide-generating NADPH oxidase activator, sphingomyelin phosphodiesterase activator, monophenol monooxygenase activator, proteasome activator, and GTPase activator can be used to diagnose/treat diseases in which beneficial effect may be achieved by modulating the activity of activators of proteins and enzymes.

Cell growth and/or maintenance proteins

Information derived from proteins involved in any biological process required for cell survival, growth and maintenance including proteins involved in cell organization and biogenesis, cell growth, cell proliferation, metabolism, cell cycle, budding, cell shape and cell size control, sporulation (sensu Saccharomyces), transport, ion homeostasis, autophagy, cell motility, chemi-mechanical coupling, membrane fusion, cell-cell fusion and stress response can be used to diagnose/treat diseases involving premature death of cells, such as degenerative diseases, for example neurodegenerative diseases or conditions

associated with aging, or alternatively, diseases in which cell apoptosis is not turned on, such as cancerous diseases.

Metabolic proteins

Information derived from proteins involved in carbohydrate metabolism, energy pathways, electron transport, nucleobase, nucleoside, nucleotide and nucleic acid metabolism, protein metabolism and modification, amino acid and derivative metabolism, protein targeting, lipid metabolism, aromatic compound metabolism, one-carbon compound metabolism, coenzymes and prosthetic group metabolism, sulfur metabolism, phosphorus metabolism, phosphate metabolism, oxygen and radical metabolism, xenobiotic metabolism, nitrogen metabolism, fat body metabolism (sensu Insecta), protein localization, catabolism, biosynthesis, toxin metabolism , methylglyoxal metabolism, cyanate metabolism, glycolate metabolism, carbon utilization, and antibiotic metabolism can be used to treat or detect diseases in which metabolism of small molecules and macromolecules such as toxins, lipids, proteins and carbohydrates is abnormal leading to various pathologies.

Channel/pore class transporters

Information derived from proteins that mediate the transport of molecules and macromoleules across membranes, such as alpha-type channels, porins and pore-forming toxins can be used to diagnose/treat diseases in which the transport of molecules and macromolecules such as neurotransmitters, hormones, sugar etc. is non-normal leading to various pathologies.

Tubulin binding proteins

Information derived from proteins that bind tubulin, such as microtubule binding proteins can be used to diagnose/treat diseases involving abnormal tubulin activity or structure. Binding of the RNA products of the genes of this family, or antibodies reactive therewith, can modulate a plurality of tubulin activities as well as change microtubulin structure.

Kinases

Information derived from kinases such as 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase, NAD(+) kinase, acetylglutamate kinase, adenosine kinase, adenylate kinase, adenylsulfate kinase, arginine kinase, aspartate kinase, choline kinase, creatine kinase, cytidylate kinase, deoxyadenosine kinase, deoxycytidine kinase, deoxyguanosine kinase, dephospho-CoA kinase, diacylglycerol kinase, dolichol kinase, ethanolamine kinase, galactokinase, glucokinase, glutamate 5-kinase, glycerol kinase, glycerone kinase, guanylate kinase, hexokinase, homoserine kinase, hydroxyethylthiazole kinase, inositol/phosphatidylinositol kinase, ketohexokinase, mevalonate kinase, nucleoside-diphosphate kinase, pantothenate kinase, phosphoenolpyruvate carboxykinase, phosphoglycerate kinase, phosphomevalonate kinase, protein kinase, pyruvate dehydrogenase (lipoamide) kinase, pyruvate kinase, ribokinase, ribose-phosphate pyrophosphokinase, selenide,water dikinase, shikimate kinase, thiamine pyrophosphokinase, thymidine kinase, thymidylate kinase, uridine kinase, xylulokinase, 1D-myo-inositol-trisphosphate 3-kinase, phosphofructokinase, pyridoxal kinase, sphinganine kinase, riboflavin kinase, 2-dehydro-3-deoxygalactonokinase, 2-dehydro-3-deoxygluconokinase, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, GTP pyrophosphokinase, L-fuculokinase, L-ribulokinase, L-xylulokinase, isocitrate dehydrogenase (NADP+)] kinase, acetate kinase, allose kinase, carbamate kinase, cobinamide kinase, diphosphate-purine nucleoside kinase, fructokinase, glycerate kinase, hydroxymethylpyrimidine kinase, hygromycin-B kinase, inosine kinase, kanamycin kinase, phosphomethylpyrimidine kinase, phosphoribulokinase, polyphosphate kinase, propionate kinase, pyruvate,water dikinase, rhamnulokinase, tagatose-6-phosphate kinase, tetraacyldisaccharide 4'-kinase, thiamine-phosphate kinase, undecaprenol kinase, uridylate kinase, N-acylmannosamine kinase and D-erythro-sphingosine kinase can be used to

diagnose/treat diseases, which may be ameliorated by a modulating kinase activity, which is one of the main signaling pathways inside cells.

Oxidoreductases

Information derived from enzymes that catalyze an oxidation-reduction reaction, including oxidoreductases acting on CH-OH, CH-CH, CH-NH2, CH-NH, NADH or NADPH, nitrogenous compounds, sulfur group of donors, heme group, hydrogen group, diphenols and related substances as donors, oxidoreductases acting on peroxide as acceptor, superoxide radicals as acceptor, oxidizing metal ions, CH2 groups, reduced ferredoxin donor, reduced flavodoxin donor, and aldehyde or oxo group of donors can be used to diagnose/treat diseases involving non-normal activity of oxidoreductases.

Transferases

Information derived from enzymes that catalyze the transfer of a chemical group, such as a phosphate or amine, from one molecule to another including transferases, transferring one-carbon groups, aldehyde or ketonic groups, acyl groups, glycosyl groups, alkyl or aryl (other than methyl) groups, nitrogenous, phosphorus-containing groups, sulfur-containing groups and lipoyltransferase, deoxycytidyl transferases can be used to diagnose/treat diseases in which the transfer of a chemical group from one molecule to another is abnormal and a beneficial effect may be achieved by modulation of such abnormal reactions.

Transferases - one-carbon group

Information derived from enzymes that catalyze the transfer of a single carbon from one molecule to another including methyltransferase, amidinotransferase, hydroxymethyl-, formyl- and related transferase, carboxyl- and carbamoyltransferase can be used to diagnose/treat diseases in which the transfer of a one-carbon chemical group from one molecule to another is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Transferases - glycosyl groups

Information derived from enzymes that catalyze the transfer of a glycosyl from one molecule to another including murein lytic endotransglycosylase E and sialyltransferase can be used to diagnose/treat diseases in which the transfer of a glycosyl chemical group from one molecule to another is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Transferases - phosphorus-containing groups

Information derived from enzymes that catalyze the transfer of phosphate from one molecule to another can be used to diagnose/treat diseases in which the transfer of a phosphate group to a modulated moiety is abnormal and a beneficial effect may be achieved by modulation of such abnormal transfer.

Hydrolases

Information derived from hydrolytic enzymes acting on ester bonds, glycosyl bonds, ether bonds, carbon-nitrogen (but not peptide) bonds, acid anhydrides, acid carbon-carbon bonds, acid halide bonds, acid phosphorus-nitrogen bonds, acid sulfur-nitrogen bonds, acid carbon-phosphorus bonds and acid sulfur-sulfur bonds can be used to diagnose/treat diseases in which the hydrolytic cleavage of a covalent bond with accompanying addition of water, -H being added to one product of the cleavage and –OH to the other, is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Hydrolases, acting on ester bonds

Information derived from hydrolytic enzymes, acting on ester bonds, such as nucleases, sulfuric ester hydrolase, carboxylic ester hydrolases, thiolester hydrolase, phosphoric monoester hydrolase, phosphoric diester hydrolase, triphosphoric monoester hydrolase, diphosphoric monoester hydrolase and phosphoric triester hydrolase can be used to diagnose/treat diseases in which the hydrolytic cleavage of a covalent bond with

accompanying addition of water, -H being added to one product of the cleavage and –OH to the other, is abnormal and a beneficial effect may be achieved by modulation of such an abnormalreaction.

Carboxylic ester hydrolases

Information derived from hydrolytic enzymes, acting on carboxylic ester bonds, such as N-acetylglucosaminylphosphatidylinositol deacetylase, 2-acetyl-1-alkylglycerophosphocholine esterase, aminoacyl-tRNA hydrolase, arylesterase, carboxylesterase, cholinesterase, gluconolactonase, sterol esterase, acetylesterase, carboxymethylenebutenolidase, protein-glutamate methylesterase, and lipase, 6-phosphogluconolactonase can be used to diagnose/treat diseases which the hydrolytic cleavage of a covalent bond with accompanying addition of water, -H being added to one product of the cleavage and –OH to the other, is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Phosphoric monoester hydrolases

Information derived from hydrolytic enzymes acting on ester bonds, such as nuclease, sulfuric ester hydrolase, carboxylic ester hydrolase, thiolester hydrolase, phosphoric monoester hydrolase, phosphoric diester hydrolase, triphosphoric monoester hydrolase, diphosphoric monoester hydrolase and phosphoric triester hydrolase can be used to diagnose/treat diseases in which the hydrolytic cleavage of a covalent bond with accompanying addition of water, -H being added to one product of the cleavage and –OH to the other, is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Hydrolases acting on glycosyl bonds

Information derived from hydrolytic enzymes that act on glycosyl bonds, such as hydrolases hydrolyzing N-glycosyl compounds and S-glycosyl compounds, O-glycosyl compounds can be used to diagnose/treat diseases in which the hydrolase-related activities are abnormal.

Hydrolases acting on acid anhydrides

Information derived from hydrolytic enzymes which act on acid anhydrides, such as phosphorus-containing anhydrides, sulfonyl-containing anhydrides, and hydrolases catalysing transmembrane movement of substances, and involved in cellular and subcellular movement can be used to diagnose/treat diseases in which the hydrolase-related activities are abnormal.

Lyases

Information derived from enzymes that catalyze the formation of double bonds by removing chemical groups from a substrate without hydrolysis or catalyze the addition of chemical groups to double bonds including carbon-carbon lyases, carbon-oxygen lyases, carbon-nitrogen lyases, carbon-sulfur lyases, carbon-halide lyases, phosphorus-oxygen lyases, and other lyases can be used to diagnose/treat diseases in which lyase activity, expression or localization is abnormal.

Ligases

Information derived from enzymes that catalyze the linkage of two molecules, generally utilizing ATP as the energy donor can be used to diagnose/treat diseases in which the joining together of two molecules in an energy-dependent process is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

Ligases catalyzing carbon-oxygen bonds

Information derived from enzymes that catalyze the linkage between carbon and oxygen, such as ligase forming aminoacyl-tRNA and related compounds can be used to diagnose/treat diseases in which the linkage between carbon and oxygen in an energy-dependent process is abnormal and a beneficial effect may be achieved by modulation of such an abnormal reaction.

ATPases

Information derived from enzymes such as plasma membrane cation-transporting ATPase, ATP-binding cassette (ABC) transporter, magnesium-ATPase, hydrogen-/sodium-translocating ATPase, arsenite-transporting

ATPase, protein-transporting ATPase, DNA translocase, and P-type ATPase can be used to diagnose/treat diseases associated with abnormal activity of an ATP hydrolyzing enzyme.

ii. Diseases

Various types of diseases can be diagnosed/treated using the teachings of the present invention.

Inflammatory diseases

Examples of inflamatory diseases Include, but are not limited to, chronic inflammatory diseases and acute inflammatory diseases.

Inflammatory diseases associated with hypersensitivity

Examples of hypersensitivity include, but are not limited to, Types I-IV hypersensitivity, immediate hypersensitivity, antibody mediated hypersensitivity, immune complex mediated hypersensitivity, T lymphocyte mediated hypersensitivity and DTH.

An example of type I or immediate hypersensitivity is asthma. Examples of type II hypersensitivity include, but are not limited to, rheumatoid diseases, rheumatoid autoimmune diseases, rheumatoid arthritis (Krenn V. *et al.,* Histol Histopathol 2000 Jul;15 (3):791), spondylitis, ankylosing spondylitis (Jan Voswinkel *et al.,* Arthritis Res 2001; 3 (3): 189), systemic diseases, systemic autoimmune diseases, systemic lupus erythematosus (Erikson J. *et al.,* Immunol Res 1998;17 (1-2):49), sclerosis, systemic sclerosis (Renaudineau Y. *et al.,* Clin Diagn Lab Immunol. 1999 Mar;6 (2):156); Chan OT. *et al.,* Immunol Rev 1999 Jun;169:107), glandular diseases, glandular autoimmune diseases, pancreatic autoimmune diseases, diabetes, Type I diabetes (Zimmet P. Diabetes Res Clin Pract 1996 Oct;34 Suppl:S125), thyroid diseases, autoimmune thyroid diseases, Graves' disease (Orgiazzi J. Endocrinol Metab Clin North Am 2000 Jun;29 (2):339), thyroiditis, spontaneous autoimmune thyroiditis (Braley-Mullen H. and Yu S, J Immunol 2000 Dec 15;165 (12):7262), Hashimoto's thyroiditis (Toyoda N. *et al.,* Nippon Rinsho 1999 Aug;57 (8):1810), myxedema, idiopathic myxedema (Mitsuma T. Nippon Rinsho. 1999 Aug;57 (8):1759);

autoimmune reproductive diseases, ovarian diseases, ovarian autoimmunity (Garza KM. *et al.,* J Reprod Immunol 1998 Feb;37 (2):87), autoimmune anti-sperm infertility (Diekman AB. *et al.,* Am J Reprod Immunol. 2000 Mar;43 (3):134), repeated fetal loss (Tincani A. *et al.,* Lupus 1998;7 Suppl 2:S107-9), neurodegenerative diseases, neurological diseases, neurological autoimmune diseases, multiple sclerosis (Cross AH. *et al.*, J Neuroimmunol 2001 Jan 1;112 (1-2):1), Alzheimer's disease (Oron L. *et al.,* J Neural Transm Suppl. 1997;49:77), myasthenia gravis (Infante AJ. And Kraig E, Int Rev Immunol 1999;18 (1-2):83), motor neuropathies (Kornberg AJ. J Clin Neurosci. 2000 May;7 (3):191), Guillain-Barre syndrome, neuropathies and autoimmune neuropathies (Kusunoki S. Am J Med Sci. 2000 Apr;319 (4):234), myasthenic diseases, Lambert-Eaton myasthenic syndrome (Takamori M. Am J Med Sci. 2000 Apr;319 (4):204), paraneoplastic neurological diseases, cerebellar atrophy, paraneoplastic cerebellar atrophy, non-paraneoplastic stiff man syndrome, cerebellar atrophies, progressive cerebellar atrophies, encephalitis, Rasmussen's encephalitis, amyotrophic lateral sclerosis, Sydeham chorea, Gilles de la Tourette syndrome, polyendocrinopathies, autoimmune polyendocrinopathies (Antoine JC. and Honnorat J. Rev Neurol (Paris) 2000 Jan;156 (1):23); neuropathies, dysimmune neuropathies (Nobile-Orazio E. *et al.*, Electroencephalogr Clin Neurophysiol Suppl 1999;50:419); neuromyotonia, acquired neuromyotonia, arthrogryposis multiplex congenita (Vincent A. *et al.,* Ann N Y Acad Sci. 1998 May 13;841:482), cardiovascular diseases, cardiovascular autoimmune diseases, atherosclerosis (Matsuura E. *et al.,* Lupus. 1998;7 Suppl 2:S135), myocardial infarction (Vaarala O. Lupus. 1998;7 Suppl 2:S132), thrombosis (Tincani A. *et al.,* Lupus 1998;7 Suppl 2:S107-9), granulomatosis, Wegener's granulomatosis, arteritis, Takayasu's arteritis and Kawasaki syndrome (Praprotnik S. *et al.,* Wien Klin Wochenschr 2000 Aug 25;112 (15-16):660); anti-factor VIII autoimmune disease (Lacroix-Desmazes S. *et al.,* Semin Thromb Hemost.2000;26 (2):157); vasculitises, necrotizing small vessel vasculitises, microscopic polyangiitis, Churg and Strauss

syndrome, glomerulonephritis, pauci-immune focal necrotizing glomerulonephritis, crescentic glomerulonephritis (Noel LH. Ann Med Interne (Paris). 2000 May;151 (3):178); antiphospholipid syndrome (Flamholz R. *et al.,* J Clin Apheresis 1999;14 (4):171); heart failure, agonist-like beta-adrenoceptor antibodies in heart failure (Wallukat G. *et al.,* Am J Cardiol. 1999 Jun 17;83 (12A):75H), thrombocytopenic purpura (Moccia F. Ann Ital Med Int. 1999 Apr-Jun;14 (2):114); hemolytic anemia, autoimmune hemolytic anemia (Efremov DG. *et al.,* Leuk Lymphoma 1998 Jan;28 (3-4):285), gastrointestinal diseases, autoimmune diseases of the gastrointestinal tract, intestinal diseases, chronic inflammatory intestinal disease (Garcia Herola A. *et al.,* Gastroenterol Hepatol. 2000 Jan;23 (1):16), celiac disease (Landau YE. and Shoenfeld Y. Harefuah 2000 Jan 16;138 (2):122), autoimmune diseases of the musculature, myositis, autoimmune myositis, Sjogren's syndrome (Feist E. *et al.,* Int Arch Allergy Immunol 2000 Sep;123 (1):92); smooth muscle autoimmune disease (Zauli D. *et al.,* Biomed Pharmacother 1999 Jun;53 (5-6):234), hepatic diseases, hepatic autoimmune diseases, autoimmune hepatitis (Manns MP. J Hepatol 2000 Aug;33 (2):326) and primary biliary cirrhosis (Strassburg CP. *et al.,* Eur J Gastroenterol Hepatol. 1999 Jun;11 (6):595).

Examples of type IV or T cell mediated hypersensitivity, include, but are not limited to, rheumatoid diseases, rheumatoid arthritis (Tisch R, McDevitt HO. Proc Natl Acad Sci U S A 1994 Jan 18;91 (2):437), systemic diseases, systemic autoimmune diseases, systemic lupus erythematosus (Datta SK., Lupus 1998;7 (9):591), glandular diseases, glandular autoimmune diseases, pancreatic diseases, pancreatic autoimmune diseases, Type 1 diabetes (Castano L. and Eisenbarth GS. Ann. Rev. Immunol. 8:647); thyroid diseases, autoimmune thyroid diseases, Graves' disease (Sakata S. *et al.,* Mol Cell Endocrinol 1993 Mar;92 (1):77); ovarian diseases (Garza KM. *et al.,* J Reprod Immunol 1998 Feb;37 (2):87), prostatitis, autoimmune prostatitis (Alexander RB. *et al.,* Urology 1997 Dec;50 (6):893), polyglandular syndrome, autoimmune polyglandular syndrome, Type I autoimmune polyglandular

syndrome (Hara T. *et al.,* Blood. 1991 Mar 1;77 (5):1127), neurological diseases, autoimmune neurological diseases, multiple sclerosis, neuritis, optic neuritis (Soderstrom M. *et al.,* J Neurol Neurosurg Psychiatry 1994 May;57 (5):544), myasthenia gravis (Oshima M. *et al.,* Eur J Immunol 1990 Dec;20 (12):2563), stiff-man syndrome (Hiemstra HS. *et al.,* Proc Natl Acad Sci U S A 2001 Mar 27;98 (7):3988), cardiovascular diseases, cardiac autoimmunity in Chagas' disease (Cunha-Neto E. *et al.,* J Clin Invest 1996 Oct 15;98 (8):1709), autoimmune thrombocytopenic purpura (Semple JW. *et al.,* Blood 1996 May 15;87 (10):4245), anti-helper T lymphocyte autoimmunity (Caporossi AP. *et al.,* Viral Immunol 1998;11 (1):9), hemolytic anemia (Sallah S. *et al.,* Ann Hematol 1997 Mar;74 (3):139), hepatic diseases, hepatic autoimmune diseases, hepatitis, chronic active hepatitis (Franco A. *et al.,* Clin Immunol Immunopathol 1990 Mar;54 (3):382), biliary cirrhosis, primary biliary cirrhosis (Jones DE. Clin Sci (Colch) 1996 Nov;91 (5):551), nephric diseases, nephric autoimmune diseases, nephritis, interstitial nephritis (Kelly CJ. J Am Soc Nephrol 1990 Aug;1 (2):140), connective tissue diseases, ear diseases, autoimmune connective tissue diseases, autoimmune ear disease (Yoo TJ. *et al.,* Cell Immunol 1994 Aug;157 (1):249), disease of the inner ear (Gloddek B. *et al.,* Ann N Y Acad Sci 1997 Dec 29;830:266), skin diseases, cutaneous diseases, dermal diseases, bullous skin diseases, pemphigus vulgaris, bullous pemphigoid and pemphigus foliaceus.

Examples of delayed type hypersensitivity include, but are not limited to, contact dermatitis and drug eruption.

Autoimmune diseases

Examples of autoimmune diseases include, but are not limited to, cardiovascular diseases, rheumatoid diseases, glandular diseases, gastrointestinal diseases, cutaneous diseases, hepatic diseases, neurological diseases, muscular diseases, nephric diseases, diseases related to reproduction, connective tissue diseases and systemic diseases.

Examples of autoimmune cardiovascular diseases include, but are not limited to atherosclerosis (Matsuura E. *et al.,* Lupus. 1998;7 Suppl 2:S135), myocardial infarction (Vaarala O. Lupus. 1998;7 Suppl 2:S132), thrombosis (Tincani A. *et al.,* Lupus 1998;7 Suppl 2:S107-9), Wegener's granulomatosis, Takayasu's arteritis, Kawasaki syndrome (Praprotnik S. *et al.,* Wien Klin Wochenschr 2000 Aug 25;112 (15-16):660), anti-factor VIII autoimmune disease (Lacroix-Desmazes S. *et al.,* Semin Thromb Hemost.2000;26 (2):157), necrotizing small vessel vasculitis, microscopic polyangiitis, Churg and Strauss syndrome, pauci-immune focal necrotizing and crescentic glomerulonephritis (Noel LH. Ann Med Interne (Paris). 2000 May;151 (3):178), antiphospholipid syndrome (Flamholz R. *et al.,* J Clin Apheresis 1999;14 (4):171), antibody-induced heart failure (Wallukat G. *et al.,* Am J Cardiol. 1999 Jun 17;83 (12A):75H), thrombocytopenic purpura (Moccia F. Ann Ital Med Int. 1999 Apr-Jun;14 (2):114; Semple JW. *et al.,* Blood 1996 May 15;87 (10):4245), autoimmune hemolytic anemia (Efremov DG. *et al.,* Leuk Lymphoma 1998 Jan;28 (3-4):285; Sallah S. *et al.,* Ann Hematol 1997 Mar;74 (3):139), cardiac autoimmunity in Chagas' disease (Cunha-Neto E. *et al.,* J Clin Invest 1996 Oct 15;98 (8):1709) and anti-helper T lymphocyte autoimmunity (Caporossi AP. *et al.,* Viral Immunol 1998;11 (1):9).

Examples of autoimmune rheumatoid diseases include, but are not limited to rheumatoid arthritis (Krenn V. *et al.,* Histol Histopathol 2000 Jul;15 (3):791; Tisch R, McDevitt HO. Proc Natl Acad Sci units S A 1994 Jan 18;91 (2):437) and ankylosing spondylitis (Jan Voswinkel *et al.,* Arthritis Res 2001; 3 (3): 189).

Examples of autoimmune glandular diseases include, but are not limited to, pancreatic disease, Type I diabetes, thyroid disease, Graves' disease, thyroiditis, spontaneous autoimmune thyroiditis, Hashimoto's thyroiditis, idiopathic myxedema, ovarian autoimmunity, autoimmune anti-sperm infertility, autoimmune prostatitis and Type I autoimmune polyglandular syndrome. diseases include, but are not limited to autoimmune diseases of the

pancreas, Type 1 diabetes (Castano L. and Eisenbarth GS. Ann. Rev. Immunol. 8:647; Zimmet P. Diabetes Res Clin Pract 1996 Oct;34 Suppl:S125), autoimmune thyroid diseases, Graves' disease (Orgiazzi J. Endocrinol Metab Clin North Am 2000 Jun;29 (2):339; Sakata S. *et al.*, Mol Cell Endocrinol 1993 Mar;92 (1):77), spontaneous autoimmune thyroiditis (Braley-Mullen H. and Yu S, J Immunol 2000 Dec 15;165 (12):7262), Hashimoto's thyroiditis (Toyoda N. *et al.*, Nippon Rinsho 1999 Aug;57 (8):1810), idiopathic myxedema (Mitsuma T. Nippon Rinsho. 1999 Aug;57 (8):1759), ovarian autoimmunity (Garza KM. *et al.*, J Reprod Immunol 1998 Feb;37 (2):87), autoimmune anti-sperm infertility (Diekman AB. *et al.*, Am J Reprod Immunol. 2000 Mar;43 (3):134), autoimmune prostatitis (Alexander RB. *et al.*, Urology 1997 Dec;50 (6):893) and Type I autoimmune polyglandular syndrome (Hara T. *et al.*, Blood. 1991 Mar 1;77 (5):1127).

Examples of autoimmune gastrointestinal diseases include, but are not limited to, chronic inflammatory intestinal diseases (Garcia Herola A. *et al.*, Gastroenterol Hepatol. 2000 Jan;23 (1):16), celiac disease (Landau YE. and Shoenfeld Y. Harefuah 2000 Jan 16;138 (2):122), colitis, ileitis and Crohn's disease.

Examples of autoimmune cutaneous diseases include, but are not limited to, autoimmune bullous skin diseases, such as, but are not limited to, pemphigus vulgaris, bullous pemphigoid and pemphigus foliaceus.

Examples of autoimmune hepatic diseases include, but are not limited to, hepatitis, autoimmune chronic active hepatitis (Franco A. *et al.*, Clin Immunol Immunopathol 1990 Mar;54 (3):382), primary biliary cirrhosis (Jones DE. Clin Sci (Colch) 1996 Nov;91 (5):551; Strassburg CP. *et al.*, Eur J Gastroenterol Hepatol. 1999 Jun;11 (6):595) and autoimmune hepatitis (Manns MP. J Hepatol 2000 Aug;33 (2):326).

Examples of autoimmune neurological diseases include, but are not limited to, multiple sclerosis (Cross AH. *et al.*, J Neuroimmunol 2001 Jan 1;112 (1-2):1), Alzheimer's disease (Oron L. *et al.*, J Neural Transm Suppl.

1997;49:77), myasthenia gravis (Infante AJ. And Kraig E, Int Rev Immunol 1999;18 (1-2):83; Oshima M. *et al.,* Eur J Immunol 1990 Dec;20 (12):2563), neuropathies, motor neuropathies (Kornberg AJ. J Clin Neurosci. 2000 May;7 (3):191); Guillain-Barre syndrome and autoimmune neuropathies (Kusunoki S. Am J Med Sci. 2000 Apr;319 (4):234), myasthenia, Lambert-Eaton myasthenic syndrome (Takamori M. Am J Med Sci. 2000 Apr;319 (4):204); paraneoplastic neurological diseases, cerebellar atrophy, paraneoplastic cerebellar atrophy and stiff-man syndrome (Hiemstra HS. *et al.,* Proc Natl Acad Sci units S A 2001 Mar 27;98 (7):3988); non-paraneoplastic stiff man syndrome, progressive cerebellar atrophies, encephalitis, Rasmussen's encephalitis, amyotrophic lateral sclerosis, Sydeham chorea, Gilles de la Tourette syndrome and autoimmune polyendocrinopathies (Antoine JC. and Honnorat J. Rev Neurol (Paris) 2000 Jan;156 (1):23); dysimmune neuropathies (Nobile-Orazio E. *et al.,* Electroencephalogr Clin Neurophysiol Suppl 1999;50:419); acquired neuromyotonia, arthrogryposis multiplex congenita (Vincent A. *et al.,* Ann N Y Acad Sci. 1998 May 13;841:482), neuritis, optic neuritis (Soderstrom M. *et al.,* J Neurol Neurosurg Psychiatry 1994 May;57 (5):544) and neurodegenerative diseases.

Examples of autoimmune muscular diseases include, but are not limited to, myositis, autoimmune myositis and primary Sjogren's syndrome (Feist E. *et al.,* Int Arch Allergy Immunol 2000 Sep;123 (1):92) and smooth muscle autoimmune disease (Zauli D. *et al.,* Biomed Pharmacother 1999 Jun;53 (5-6):234).

Examples of autoimmune nephric diseases include, but are not limited to, nephritis and autoimmune interstitial nephritis (Kelly CJ. J Am Soc Nephrol 1990 Aug;1 (2):140).

Examples of autoimmune diseases related to reproduction include, but are not limited to, repeated fetal loss (Tincani A. *et al.,* Lupus 1998;7 Suppl 2:S107-9).

Examples of autoimmune connective tissue diseases include, but are not limited to, ear diseases, autoimmune ear diseases (Yoo TJ. *et al.*, Cell Immunol 1994 Aug;157 (1):249) and autoimmune diseases of the inner ear (Gloddek B. *et al.*, Ann N Y Acad Sci 1997 Dec 29;830:266).

Examples of autoimmune systemic diseases include, but are not limited to, systemic lupus erythematosus (Erikson J. *et al.*, Immunol Res 1998;17 (1-2):49) and systemic sclerosis (Renaudineau Y. *et al.*, Clin Diagn Lab Immunol. 1999 Mar;6 (2):156); Chan OT. *et al.*, Immunol Rev 1999 Jun;169:107).

Infectious diseases

Examples of infectious diseases include, but are not limited to, chronic infectious diseases, subacute infectious diseases, acute infectious diseases, viral diseases, bacterial diseases, protozoan diseases, parasitic diseases, fungal diseases, mycoplasma diseases and prion diseases.

Graft rejection diseases

Examples of diseases associated with transplantation of a graft include, but are not limited to, graft rejection, chronic graft rejection, subacute graft rejection, hyperacute graft rejection, acute graft rejection and graft versus host disease.

Allergic diseases

Examples of allergic diseases include, but are not limited to, asthma, hives, urticaria, pollen allergy, dust mite allergy, venom allergy, cosmetics allergy, latex allergy, chemical allergy, drug allergy, insect bite allergy, animal dander allergy, stinging plant allergy, poison ivy allergy and food allergy.

Cancerous diseases

Examples of cancer include but are not limited to carcinoma, lymphoma, blastoma, sarcoma, and leukemia. Particular examples of cancerous diseases but are not limited to: Myeloid leukemia such as Chronic myelogenous leukemia. Acute myelogenous leukemia with maturation. Acute promyelocytic leukemia, Acute nonlymphocytic leukemia with increased basophils, Acute monocytic leukemia. Acute myelomonocytic leukemia with eosinophilia;

Malignant lymphoma, such as Birkitt's Non-Hodgkin's; Lymphoctyic leukemia, such as Acute lumphoblastic leukemia. Chronic lymphocytic leukemia; Myeloproliferative diseases, such as Solid tumors Benign Meningioma, Mixed tumors of salivary gland, Colonic adenomas; Adenocarcinomas, such as Small cell lung cancer, Kidney, Uterus, Prostate, Bladder, Ovary, Colon, Sarcomas, Liposarcoma, myxoid, Synovial sarcoma, Rhabdomyosarcoma (alveolar), Extraskeletel myxoid chonodrosarcoma, Ewing's tumor; other include Testicular and ovarian dysgerminoma, Retinoblastoma, Wilms' tumor, Neuroblastoma, Malignant melanoma, Mesothelioma, breast, skin, prostate, and ovarian.

EXAMPLE 9

Microarray analysis based validation of the antisense dataset

A microarray-based analysis using oligonucleotide probes that hybridize to the target in a strand-specific manner, was conducted in order to experimentally validate the predicted antisense/sense pairs of the database. Two complementary 60-mer oligonucleotide probes derived from the predicted overlap region of the sense/antisense pairs, were designed. Single 60-mer oligonucleotides were previously shown to offer reliability and sensitivity for detecting specific transcripts (T. R. Hughes, *et al.*, *Nature Biotech.* **19**, 342 (2001).).Initially only pairs of clusters with an overlap greater than 60 bases (2,464 pairs agree with this restriction) were selected for array construction. The overlap region of each antisense pair was then verified for the presence of 60-mer oligonucleotides that matched a set of standards, such as minimal sequence similarity elsewhere in the human genome, uniform GC-content and Tm, and absence of palindromic sequences, in order to maximize the hybridization specificity. Oligonucleotide probes meeting the criteria set forth were identified for 1,211 sense/antisense pairs and a random sample of 264 pairs, which constitutes roughly one-tenth of the original dataset of 2667 sense/antisense cluster pairs, was selected for analysis by Microarrays (Table

S1 on CD-ROM3, an excerpt of which is shown in Table 5 below). In this sample, the proportion of each of the nine subgroups depicted in Table 4 is similar to that of the original dataset, indicating a good representation of the various subgroups.

Table 4

mRNA/ Splicing	No cluster w introns	1 cluster w intron(s)	2 clusters w intron(s)	Total
No cluster w mRNA	48	132	197	377 (14%)
1 cluster w mRNA	17	490	1039	1546 (58%)
2 clusters w mRNA	1	85	658	744 (28%)
Total	66 (2.5%)	707 (26%)	1894 (71%)	2667 (100%)

Table represents the proportion of sense/antisense clusters in the dataset of 2667 that contain: 1) a known mRNA and 2) expressed sequences spanning at least one intron, in one of the two clusters, in both clusters or in none of the clusters.

Table 5 below is an excerpt of Table S1 provided on CD-ROM3; Table 5 exemplifies five of the putative sense/antisense pairs that were selected for microarray analysis. The first column provides the pair number. The next two columns provide the accession numbers of representative expressed sequences from the overlapping region of the sense and the antisense genes, respectively. The two columns identified by the "RNA" header provide the accession numbers of known mRNAs in the sense and antisense clusters (if available), and the last two columns provide the GenBank descriptions of these mRNAs.

Table 5

Pair no.	sense seq. from over-lapping region	antisense seq. from overlapping region	RNA in sense cluster	RNA in a-sense cluster	description of RNA in sense cluster	description of RNA in antisense cluster
235	NM_ 6227	NM_ 308	NM_ 6227	NM_ 308	Homo sapiens phospholipid transfer protein (PLTP), mRNA #DV L26232.1	Homo sapiens protective protein for beta-galactosidase (galactosialidosis) (PPGB), mRNA

237	NM_	NM_	NM_	NM_	Homo sapiens	Homo sapiens
	4703	2532	4703	2532	rabaptin-5	nucleoporin 88kD
					(RAB5EP), mRNA	(NUP88) mRNA
	.				#DV X91141.1	#DV Y08612.2
217	NM_	AV	NM_	NM_	Homo sapiens	Homo sapiens ATP-
	14885	723808	14885	2940	anaphase-promoting	binding cassette,
					complex 10	sub-family E
					(APC10) mRNA.	(OABP), member 1
	.				#DV AL080090.1	(ABCE1), mRNA.
209	BC	BG	NM_	NM_	Homo sapiens	Homo sapiens
	8865	717574	32231	3099	hypothetical protein	sorting nexin 1
					FLJ22875	(SNX1),mRNA.
					(FLJ22875), mRNA	#DV U53225.1
196	BE	AL	NM_	NM_	Homo sapiens	Homo sapiens
	885605	527611	17832	3640	hypothetical protein	inhibitor of kappa
					FLJ20457	light polypeptide gene
					(FLJ20457), mRNA	enhancer in B-cells,
						kinase complex-
						associated protein
						(IKBKAP),mRNA

Table 5 Cont.

Microarrays were constructed by spotting each of the 264 pairs of oligonucleotide probes onto treated glass slides in quadruplicates. The two counterpart oligonucleotide probes of each pair were spotted next to each other to ensure similar hybridization conditions.

As positive controls, each of the blocks contained oligonucleotides spotted at various concentrations for four ubiquitously expressed housekeeping genes: guanine nucleotide binding protein beta polypeptide 2-like 1 (gnb2l1, HUMMHBA123, NM_006098), heat shock 70kD protein 10 (hsp70, HSHSC70CDS0, NM_006597), beta actin (actin, ACTB, NM_001101), and glyceraldehyde-3-phosphate dehydrogenase (gapdh, NM_002046).

Two random oligonucleotides were used as negative controls. These computer-generated arbitrary sequences displayed no alignment to human genome sequences but had the same physical characteristics as the other

oligonucleotide probes. In addition, 22 probes for 11 previously documented sense/antisense pairs were also analyzed in the Microarrays (entries Pair no. "known 1"-"known 11" on Table S1 of CD-ROM3).

The Microarrays were hybridized with poly(A)+ RNAs obtained from 19 human cell lines representing a variety of tissues and four normal human tissues (see *General Materials and Methods* section above). Each poly(A)+ RNA was reverse transcribed by priming with oligo(dT) and random nonamers, and engineered to incorporate a fluorescent marker. A pool containing an equal mix of the RNAs from all cell lines was also transcribed and used as a reference target. The resulting fluorescently-labeled cDNAs were combined and hybridized to the oligonucleotide Microarrays.

The experiments were performed in duplicate and utilized a fluorescent reversal of the Cy3- and Cy5-labelled cDNA. Stringent hybridization conditions were utilized in order to minimize the appearance of false positive signals, despite the possibility of compromised detection of low abundance transcripts.

The raw data was normalized at several levels; within each slide, between reciprocal slides, and globally between slides (see *General Materials and Methods* section above). Non-specific levels of hybridization were estimated from the negative controls. The threshold for significant positive signals resulting from authentic hybridization was set at 4 standard deviations of the mean normalized signals for the negative controls. Processed data was presented as normalized signal intensity and as normalized signal ratios (Table S2 on CD-ROM3).

To further substantiate array results, several pairs of oligonucleotides were also utilized in Northern blot analysis. Figures 22a-j illustrate results of such northern blot analysis. Figure 22a reveals expression patterns of randomly selected sequence pair number 235, denoted as Rand_235 in Table 6. Similarly, Figure 22b corresponds to pair number 173, Figure 22c to pair number 248, Figure 22d to pair number 6, Figure 22e to pair number 216, Figure 22f to pair number 239, Figure 22g to pair number 202, Figure 22h to

pair number 114, Figure 22i to pair number 188, and Figure 22j to pair number 223. Eight pairs (Figures 22a-h) evaluated revealed positive signals for both sense and antisense expression, while two (Figures 22i-j) revealed a positive signal for only one of the genes, with the counterpart being a known RefSeq mRNA.

Figure 23 represents an excerpt of Table S2 (provided in CD-ROM3) which summarizes the results obtained utilizing the array generated according to the teachings of the present invention. Expression thresholds were verified and indicated and normalization for microarray signals was conducted as described above. *Rji* ratios were obtained for each cell line/tissue assessed.

Taken cumulatively, the data presented herein revealed positive signals for both sense and antisense transcripts in 65 cluster pairs. In another 47 cases, significant hybridization signals were detected for antisense sequences with known counterpart sense transcripts, i.e. RefSeq mRNAs, which did not give clear hybridization signals on the Microarrays. Thus, 42.5 % (112 cases) of the 264 represented on the Microarrays, yielded detectable antisense transcription.

The conversion table, assigning the respective serial number as it appears in the "Table" file of CD-ROM1 enclosed herewith, is shown in Table 6 below.

Table 6

Rand_#	Serial No	Rand_#	Serial No	Rand_#	Serial No
Rand_1	2326	Rand_179	3266	Rand_258	3807
Rand_10	3647	Rand_18	3073	Rand_259	2621
Rand_100	2758	Rand_180	1794	Rand_26	4009
Rand_101	1595	Rand_181	1585	Rand_27	3393
Rand_102	3686	Rand_182	3554	Rand_28	3589
Rand_103	2331	Rand_183	3377	Rand_29	1837
Rand_104	3496	Rand_184	3466	Rand_3	3046
Rand_105	3134	Rand_185	3159	Rand_30	3297
Rand_106	1339	Rand_186	1413	Rand_31	3692
Rand_107	908	Rand_187	3645	Rand_32	707 2376
Rand_108	2929	Rand_188	3880	Rand_33	2052

Rand_109	2537	Rand_189	3009	Rand_34	1904
Rand_11	2806	Rand_19	3641	Rand_35	3718
Rand_110	3594	Rand_190	2549	Rand_36	3898
Rand_111	2819	Rand_191	2874	Rand_37	1821
Rand_112	3019	Rand_192	2515	Rand_38	3092
Rand_113	3815	Rand_193	3914	Rand_39	3262
Rand_114	2606	Rand_194	2751	Rand_4	3558
Rand_115	1662	Rand_195	2091	Rand_40	2474
Rand_116	2171	Rand_196	1966	Rand_41	3568
Rand_117	2539	Rand_197	3778	Rand_42	864
Rand_118	2802	Rand_198	3877	Rand_43	1864
Rand_119	2761	Rand_199	2248	Rand_44	3045
Rand_12	1947	Rand_2	3172	Rand_45	2854
Rand_120	3228	Rand_20	2360	Rand_46	3852
Rand_121	2076	Rand_200	2064	Rand_47	3096
Rand_122	1835	Rand_201	3597	Rand_48	1987
Rand_123	3029	Rand_202	2826	Rand_49	2893
Rand_124	2898	Rand_203	2388	Rand_5	2060
Rand_125	1568	Rand_204	3889	Rand_50	1058
Rand_126	2456	Rand_205	2211	Rand_51	3560
Rand_127	2019	Rand_206	3512	Rand_52	2604
Rand_128	2346	Rand_207	3452	Rand_53	3397
Rand_129	2460	Rand_208	3886	Rand_54	2040
Rand_13	2429	Rand_209	1600	Rand_55	3784
Rand_130	3374	Rand_21	2952	Rand_56	3659
Rand_131	3292	Rand_210	2432	Rand_57	2005 2688
Rand_132	3259	Rand_211	1651 3968	Rand_58	3187
Rand_133	3591	Rand_212	3074	Rand_59	1350
Rand_134	3340	Rand_213	2341	Rand_6	2202
Rand_135	1958	Rand_214	1984	Rand_60	3183
Rand_136	2274	Rand_215	2803	Rand_61	2275
Rand_137	3527	Rand_216	3806	Rand_62	3882
Rand_138	1533	Rand_217	2186	Rand_63	1044 3899
Rand_139	2622	Rand_218	857	Rand_64	2811
Rand_14	2058	Rand_219	1744	Rand_65	3232
Rand_140	2578	Rand_22	2285	Rand_66	3242
Rand_141	3492	Rand_220	2977	Rand_67	34 112 2727

Rand_142	3928	Rand_221	3863	Rand_68	3909
Rand_143	2282 3790	Rand_222	2846	Rand_69	4016
Rand_144	2820	Rand_223	3986	Rand_7	2337
Rand_145	1329	Rand_224	579 3688	Rand_70	2101 3707
Rand_146	1783	Rand_225	3984	Rand_71	3703
Rand_147	1527	Rand_226	2889	Rand_72	3477
Rand_148	2662	Rand_227	3869	Rand_73	2437
Rand_149	2031	Rand_228	3994	Rand_74	3808
Rand_15	2677	Rand_229	3818	Rand_75	3905
Rand_150	1303 1659	Rand_23	3890	Rand_76	1138 2194
Rand_151	1767	Rand_230	3152	Rand_77	819
Rand_152	3378	Rand_231	3445	Rand_78	3704
Rand_153	984	Rand_232	3663	Rand_79	2309
Rand_154	3759	Rand_233	3410	Rand_8	3441
Rand_155	2046	Rand_234	1112	Rand_80	1219
Rand_156	2528	Rand_235	3918	Rand_81	1416
Rand_157	283 1798 2048	Rand_236	2316	Rand_82	1543
Rand_158	3710	Rand_237	3673	Rand_83	3269
Rand_159	3178	Rand_238	3990	Rand_84	532 732
Rand_16	3336	Rand_239	4012	Rand_85	2607
Rand_160	1645	Rand_24	3250	Rand_86	1867
Rand_161	2074 3464	Rand_240	2932	Rand_87	627 3006
Rand_162	3436	Rand_241	3836	Rand_88	2068
Rand_163	2738	Rand_242	3424	Rand_89	2296
Rand_164	2749	Rand_243	3982	Rand_9	3741
Rand_165	2206	Rand_244	3472	Rand_90	1076
Rand_166	1349	Rand_245	2071	Rand_91	3385
Rand_167	2773	Rand_246	3904	Rand_92	2334
Rand_168	3305	Rand_247	2056	Rand_93	2833
Rand_169	1954	Rand_248	3855	Rand_94	2626
Rand_17	3940	Rand_249	2980	Rand_95	3671
Rand_170	2813	Rand_25	3453	Rand_96	1923
Rand_171	3868	Rand_250	3565	Rand_97	1863
Rand_172	762 1424 3942	Rand_251	2459	Rand_98	3437
Rand_173	3872	Rand_252	71 3147	Rand_99	3469
Rand_174	3801	Rand_253	3967	Rand_260	1975 3171
Rand_175	2547	Rand_254	702 2867 3088	Rand_261	4013

Rand_176	1251	Rand_255	3156	Rand_262	2418
Rand_177	1603	Rand_256	2324 2998	Rand_263	2451
Rand_178	2769	Rand_257	2284	Rand_264	3832

Table 6 Cont

Rand # = the name of the pair on the chip as it appears in Table S2 on CD-ROM3, column "Probe"; Serial No = no of the pair in the Table on CD-ROM1 (could be more than one in case the antisense event was separated to more than two contigs).

The sensitivity of the experimental approach utilized, i.e. the ability to detect a given transcript, stems from a combination of the stringency used in the microarray analysis and the level of expression and tissue specificity of the RNA. This can be estimated from the positive signals obtained for 65% of the oligos representing known RefSeq mRNAs on the Microarrays. This level of detection is comparable to that obtained in other studies, such as the 58% of known exons verified using microarray analysis (D. D. Shoemaker, et al., Nature 409, 922; 2001).

Thus, the present methodology provides a level of detection for a pair of genes that is 0.65 x 0.65= 0.42, a value supported by the detection of positive signals for both sense and antisense expression in 5 out of 11 (0.45) clusters of previously described sense/antisense pairs (Table S2 on CD-ROM3).

Of the 264 cluster pairs analyzed in the Microarrays of the present invention, 65 clusters (0.25) showed significant signals for both sense and antisense transcripts, which is 60% of the proposed level of detection for a pair of genes (0.25/0.42). Extrapolating this figure to the predicted antisense dataset of 2667 clusters, predicts at least 1600 sense/antisense transcriptional units in the human genome.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications,

patents, patent applications and sequences identified by their accession numbers mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent, patent application or sequence identified by their accession number was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

CD-ROM Content

The following CD-ROMs are attached herewith:

Information provided as: File name/byte size/date of creation/operating system/machine format

CD-ROM1:

1. seqs 327MB 15/11/2001 Microsoft Windows Internet Explorer
2. table 13.5MB 15/11/2001 Microsoft Windows Internet Explorer

CD-ROM2:

1. alignments 382MB 15/11/2001 Microsoft Windows Internet Explorer

CD-ROM3:

1. Table_S1 79.5kb 10/07/2002 Microsoft Windows Microsoft Excel Worksheet
2. Table_S2 334kb 10/07/2002 Microsoft Windows Microsoft Excel Worksheet

WHAT IS CLAIMED IS:

1. A method of identifying putative naturally occurring antisense transcripts, the method comprising:

(a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b) identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database, thereby identifying putative naturally occurring antisense transcripts.

2. The method of claim 1, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

3. The method of claim 1, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

4. The method of claim 1, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

5. The method of claim 1, wherein said second database is generated by:

(i) providing a library of expressed polynucleotides;

(ii) obtaining sequence information of said expressed polynucleotides;

(iii) computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv) storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

6. The method of claim 5, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

7. The method of claim 1 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

8. The method of claim 1 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

9. A kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, said sequence region not being complementary with a naturally occurring antisense transcript.

10. The kit of claim 9, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

11. The kit of claim 9, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

12. The kit of claim 9, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

13. The kit of claim 9, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

14. The kit of claim 9, wherein said at least one oligonucleotide is labeled.

15. The kit of claim 9, wherein said at least one oligonucleotide is attached to a solid substrate.

16. The kit of claim 15, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

17. A kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one pair of oligonucleotides including a first

oligonucleotide capable of binding the at least one mRNA transcript of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest.

18. The kit of claim 17, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

19. The kit of claim 17, wherein said first and second oligonucleotides are single stranded oligonucleotides.

20. The kit of claim 17, wherein said first and second oligonucleotides are double stranded oligonucleotide.

21. The kit of claim 17, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

22. The kit of claim 17, wherein said first and second oligonucleotides are labeled.

23. The kit of claim 17, wherein said first and second oligonucleotides are attached to a solid substrate.

24. The kit of claim 23, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

25. A kit for quantifying at least one naturally occurring antisense transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the

at least one naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript.

26. The kit of claim 25, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

27. The kit of claim 25, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

28. The kit of claim 25, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

29. The kit of claim 25, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

30. The kit of claim 25, wherein said at least one oligonucleotide is labeled.

31. The kit of claim 25, wherein said at least one oligonucleotide is attached to a solid substrate.

32. The kit of claim 31, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

33. A method of designing artificial antisense transcripts, the method comprising:

(a) providing a database of naturally occurring antisense transcripts;

(b) extracting from said database criteria governing structure and/or function of said naturally occurring antisense transcripts; and

(c) designing the artificial antisense transcripts according to said criteria.

34. The method of claim 33, wherein said criteria governing structure and/or function of said naturally occurring antisense transcripts are selected from the group consisting of antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

35. The method of claim 33, wherein said step of providing said database of naturally occurring antisense transcripts is effected by:

(a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b) identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database,

(c) storing a sequence of said expressed polynucleotide sequences identified in step (b), thereby providing said database of said naturally occurring antisense transcripts..

36. The method of claim 35, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA

(cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

37. The method of claim 35, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

38. The method of claim 35, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

39. The method of claim 35, wherein said second database is generated by:

(i) providing a library of expressed polynucleotides;

(ii) obtaining sequence information of said expressed polynucleotides;

(iii) computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv) storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

40. The method of claim 39, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

41. The method of claim 35, further comprising the step of testing said putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

42. The method of claim 35 further comprising the step of computationally testing said putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

43. A computer readable storage medium comprising a database including a plurality of sequences, wherein each sequence is of a naturally occurring antisense transcript.

44. The computer readable storage medium of claim 43, wherein said database further includes information pertaining to each sequence of said naturally occurring antisense transcripts, said information is selected from the group consisting of related sense gene, antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

45. The computer readable storage medium of claim 43, wherein said database further includes information pertaining to generation of said database and potential uses of said database.

46. A method of generating a database of naturally occurring antisense transcripts, the method comprising:

(a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences;

(b) identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database so as to identify putative naturally occurring antisense transcripts; and

(c) storing sequence information of said identified naturally occurring antisense transcripts, thereby generating the database of the naturally occurring antisense transcripts.

47. The method of claim 46, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

48. The method of claim 46, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

49. The method of claim 46, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

50. The method of claim 46, wherein said second database is generated by:

(i) providing a library of expressed polynucleotides;

(ii) obtaining sequence information of said expressed polynucleotides;

(iii) computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv) storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

51. The method of claim 50, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

52. The method of claim 46 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

53. The method of claim 46 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence

information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

54. A system for generating a database of a plurality of putative naturally occurring antisense transcripts, the system comprising a processing unit, said processing unit executing a software application configured for:

(a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b) identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database.

55. The system of claim 54, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

56. The system of claim 54, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

57. The system of claim 54, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

58. The system of claim 54, wherein said second database is generated by:

(i) providing a library of expressed polynucleotides;

(ii) obtaining sequence information of said expressed polynucleotides;

(iii) computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv) storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

59. The system of claim 58, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

60. The system of claim 54 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

61. The system of claim 54 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

62. A method of identifying putative naturally occurring antisense transcripts, the method comprising screening a database of expressed polynucleotides sequences according to at least one sequence criterion, said at least one sequence criterion being selected to identify putative naturally occurring antisense transcripts.

63. The method of claim 63, wherein said database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

64. The method of claim 63, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

65. The method of claim 63, wherein said at least one sequence criterion is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

66. The method of claim 63 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form a duplex with at least one sense oriented polynucleotide sequence under physiological conditions.

67. A method of quantifying at least one mRNA of interest in a biological sample, the method comprising:

(a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one mRNA of interest,

wherein said at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, said sequence region not being complementary with a naturally occurring antisense transcript; and

(b) detecting a level of binding between the at least one mRNA of interest and said at least one oligonucleotide to thereby quantify the at least one mRNA of interest in the biological sample.

68. The method of claim 67, wherein said at least one oligonucleotide is attached to a solid substrate.

69. The method of claim 68, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

70. The method of claim 67, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

71. The method of claim 67, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

72. The method of claim 67, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

73. The method of claim 67, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

74. The method of claim 67, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

75. A method of quantifying the expression potential of at least one mRNA of interest in a biological sample, the method comprising:

(a) contacting the biological sample with at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest; and

(b) detecting a level of binding between the at least one mRNA of interest and said first oligonucleotide and a level of binding between said naturally occurring antisense transcript complementary to the mRNA of interest and said second oligonucleotide to thereby quantify the expression potential of the at least one mRNA of interest in the biological sample.

76. The method of claim 75, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

77. The method of claim 75, wherein said first and second oligonucleotides are single stranded oligonucleotides.

78. The method of claim 75, wherein said first and second oligonucleotides are double stranded oligonucleotide.

79. The method of claim 75, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

80. The method of claim 75, wherein said first and second oligonucleotides are labeled and whereas step (b) is effected by quantifying said label.

81. The method of claim 75, wherein said first and second oligonucleotides are attached to a solid substrate.

82. The method of claim 81, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

83. A method of quantifying at least one naturally occurring antisense transcript of interest in a biological sample, the method comprising:

(a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one naturally occurring antisense transcript of interest, wherein said at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript; and

(b) detecting a level of binding between the at least one naturally occurring antisense transcript of interest and said at least one oligonucleotide to thereby quantify the at least one naturally occurring antisense transcript of interest in the biological sample.

84. The method of claim 83, wherein said at least one oligonucleotide is attached to a solid substrate.

85. The method of claim 84, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

86. The method of claim 83, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

87. The method of claim 83, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

88. The method of claim 83, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

89. The method of claim 83, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

90. The method of claim 83, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

91. A method of identifying a novel drug target, the method comprising:

(a) determining expression level of at least one naturally occurring antisense transcript of interest in cells characterized by an abnormal phenotype; and

(b) comparing said expression level of said at least one naturally occurring antisense transcript of interest in said cells characterized by an abnormal phenotype to an expression level of said at least one naturally occurring antisense transcript of interest

in cells characterized by a normal phenotype, to thereby identify the novel drug target.

92. The method of claim 91, wherein said abnormal phenotype of said cells is selected from the group consisting of biochemical phenotype, morphological phenotype and nutritional phenotype.

93. The method of claim 91, wherein said determining expression level of at least one naturally occurring antisense transcript of interest is effected by at least one oligonucleotide designed and configured so as to be complementary to a sequence region of said at least one naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript.

94. The method of claim 93, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

95. The method of claim 93, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

96. The method of claim 93, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

97. The method of claim 93, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

98. The method of claim 93, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

99. The method of claim 93, wherein said at least one oligonucleotide is attached to a solid substrate.

100. The method of claim 99, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

101. A method of treating or preventing a disease, condition or syndrome associated with an upregulation of a naturally occurring antisense transcript complementary to a naturally occurring mRNA transcript, the method comprising administering a therapeutically effective amount of an agent for regulating expression of the naturally occurring antisense transcript.

102. The method of claim 101, wherein said agent for regulating expression of the naturally occurring antisense transcript is at least one oligonucleotide designed and configured so as to hybridize to a sequence region of said at least one naturally occurring antisense transcript.

103. The method of claim 102, wherein said at least one oligonucleotide is a ribozyme.

104. The method of claim 102, wherein said at least one oligonucleotide is a sense transcript.

105. A method of diagnosing a disease, condition or syndrome associated with a substandard expression ratio of an mRNA of interest over a naturally occurring antisense transcript complementary to the mRNA of interest, the method comprising:

(a) quantifying expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest;

(b) calculating the expression ratio of the mRNA of interest over the naturally occurring antisense transcript complementary to the mRNA of interest, thereby diagnosing the disease, condition or syndrome.

106. The method of claim 105, wherein quantifying said expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest is effected by at least one pair of oligonucleotides including a first oligonucleotide capable of binding the mRNA of interest and a second oligonucleotide being capable of binding the naturally occurring antisense transcript complementary to the mRNA of interest.

107. The method of claim 106, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

108. The method of claim 106, wherein said first and second oligonucleotides are single stranded oligonucleotides.

109. The method of claim 106, wherein said first and second oligonucleotides are double stranded oligonucleotides.

110. The method of claim 106, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

111. The method of claim 106, wherein said first and second oligonucleotides are labeled.

112. The method of claim 106, wherein said first and second oligonucleotides are attached to a solid substrate.

113. The method of claim 112, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

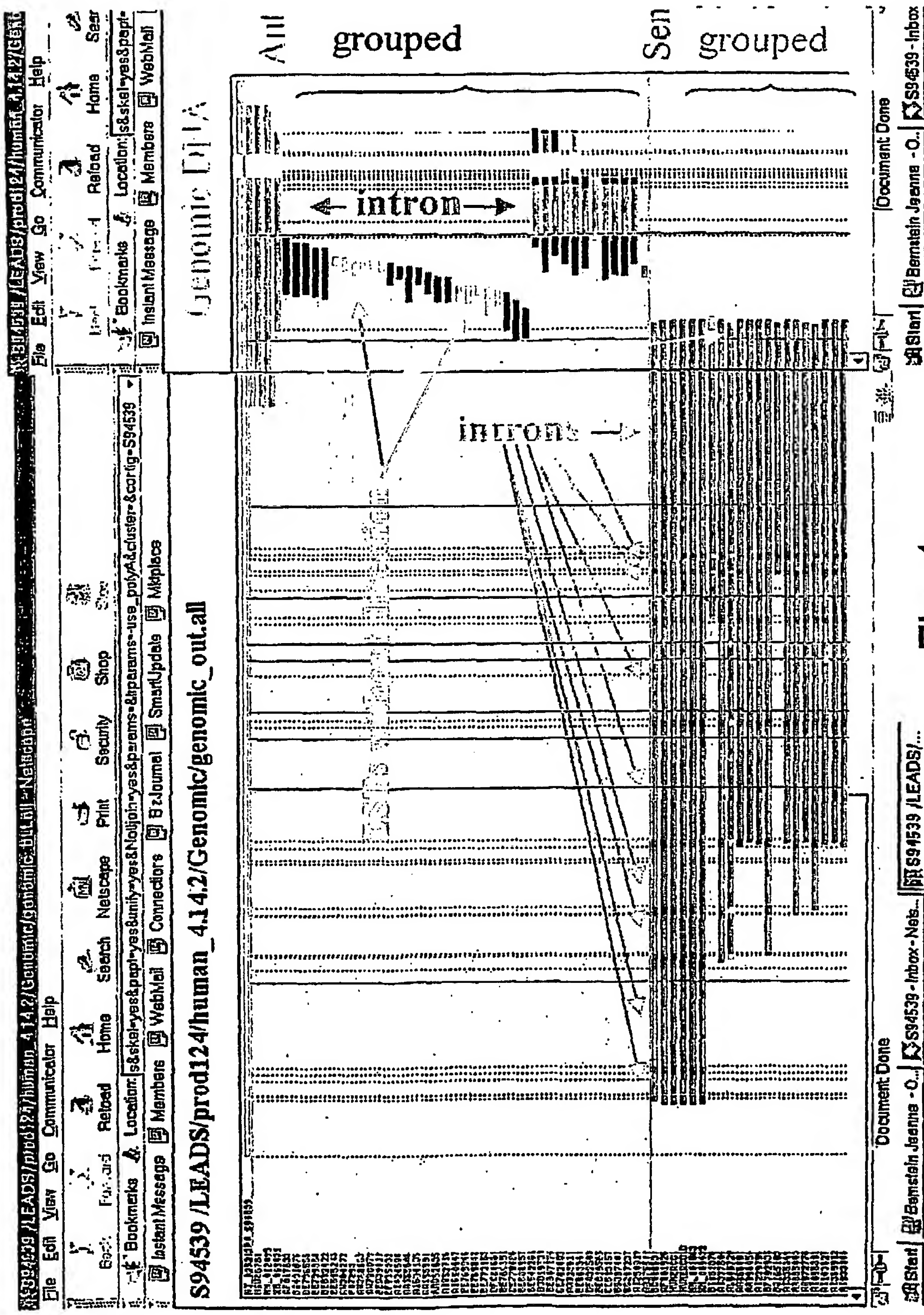

S94539 /LEADS/prod124/human_4.14.2/Genomic/genomic_out.all
All
grouped
Sen
grouped
intron
introns
Genomic DNA
ESTs
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location:
Instant Message Members WebMail Connections BizJournal SmartUpdate Mktplace
Document Done
Start Bernstein Jeanne -O... S94539-Inbox-Nets... S94539 /LEADS/...

Fig. 1

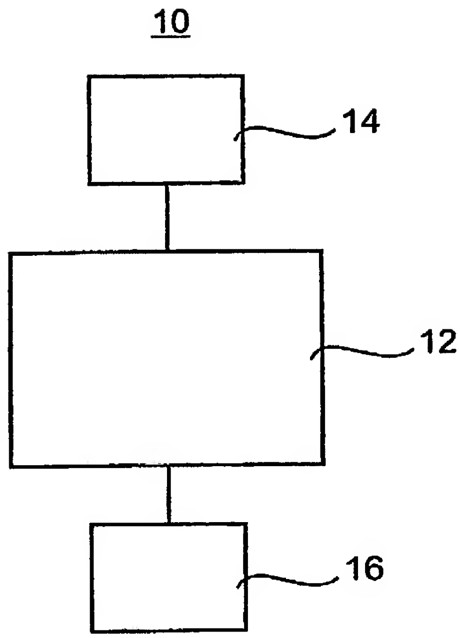


Fig. 2

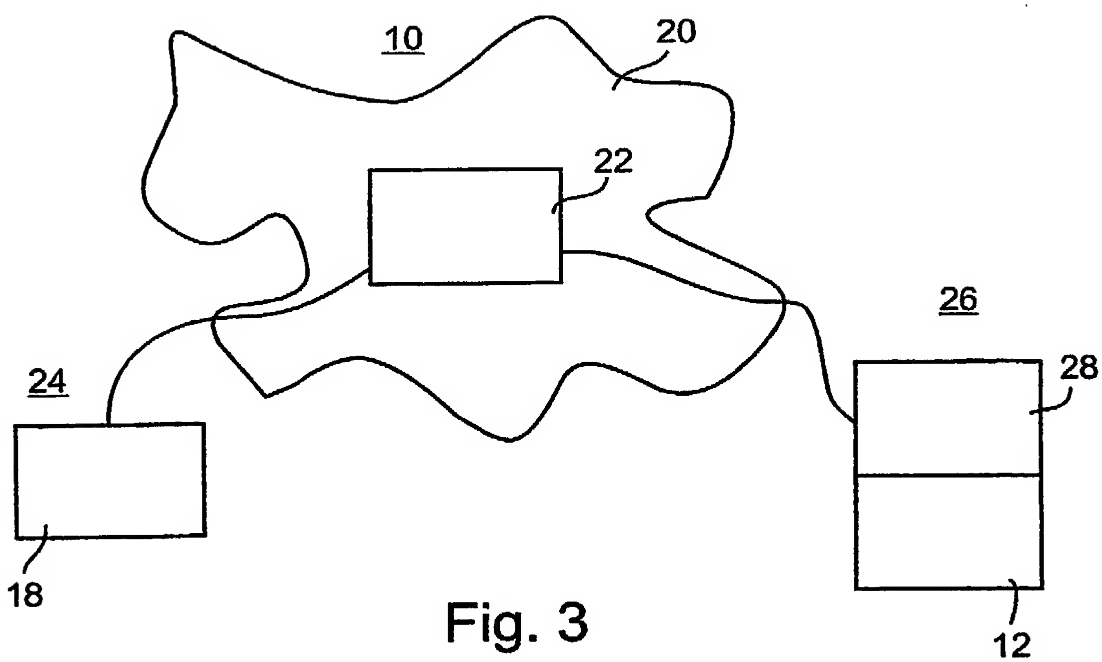


Fig. 3

Fig. 4a


```
570_0   AV705532_0      190     Z44352_15       783     OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4b

```
570_1   AV705532_0      190     Z44352_14       1649    OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4c

```
570_2   AV705532_0      190     Z44352_13       1861    OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4d

```
571_0   AW070860_0      214     T81142_7        1934    OL: 54
Query: 1    gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1215 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1162


 Score = 22.3 bits (11), Expect = 0.66
 Identities = 11/11 (100%)
 Strand = Plus / Minus


Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 760 acttccagagg 750
```

Fig. 4e

```
571_1   AW070860_0      214     T81142_6        2353    OL: 54
Query: 1    gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1215 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1162


 Score = 22.3 bits (11), Expect = 0.66
 Identities = 11/11 (100%)
 Strand = Plus / Minus


Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 760 acttccagagg 750
```

```
Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Plus


Query: 100  ggaaaacacac 110
            |||||||||||
Sbjct: 1900 ggaaaacacac 1910
```

Fig. 4f

```
571_2   AW070860_0     214     T81142_4          2500    OL: 54
Query: 1    gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1317 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1264

Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Minus


Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 862 acttccagagg 852


Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Plus


Query: 100  ggaaaacacac 110
            |||||||||||
Sbjct: 2047 ggaaaacacac 2057
```

Fig. 4g

```
571_3   AW070860_0     214     T81142_3          947     OL: 54
Query: 1   gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 224 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 171
```

Fig. 4h

```
571_4   AW070860_0     214     T81142_2          1366    OL: 54
Query: 1   gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 224 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 171


Score = 22.3 bits (11), Expect = 0.66
```

```
Identities = 11/11 (100%)
Strand = Plus / Plus

Query: 100 ggaaaacacac 110
           |||||||||||
Sbjct: 913 ggaaaacacac 923
```

Fig. 4i

```
572_0    BE046369_0      422      W26553_3         1532     OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1481 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1532
```

Fig. 4j

```
572_1    BE046369_0      422      W26553_2         1753     OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1702 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1753
```

Fig. 4k

```
572_2    BE046369_0      422      W26553_1         1832     OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1781 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1832
```

```
53BP1_76P    53BP1    10394    76P        6837     OL: 3046   OF1: 5463
OF2: 2018

 Score = 1659 bits (837), Expect = 0.0
 Identities = 840/841 (99%)
 Strand = Plus / Minus

Query: 7432 gagacggaatttcgctcttgttgcccaggctggaatgcaatggcacaatctcagctcact
7491
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3113 gagacggaatttcgctcttgttgcccaggctggaatgcaatggcacaatctcagctcact
3054

Query: 7492 gcagcgtctgcttcccaggttcaagcaattctcctgtctcagcctcctgagtagctggga
7551
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3053 gcagcgtctgcttcccaggttcaagcaattctcctgtctcagcctcctgagtagctggga
2994

Query: 7552 ttacaggcacatgccaccacacctggctaattttgtatttttagtagaatcgaggtttc
7611
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2993 ttacaggcacatgccaccacacctggctaattttgtatttttagtagaatcgaggtttc
2934

Query: 7612 atcatgttggtcaggctggtctcaaactcctgacttcaggtgatccgcccgcctcggcct
7671
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2933 atcatgttggtcaggctggtctcaaactcctgacttcaggtgatccgcccgcctcggcct
2874

Query: 7672 cccaaagtgctgggattacaggtgtgagccaccatgcccggcctaagaaatacttttaag
7731
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2873 cccaaagtgctgggattacaggtgtgagccaccatgcccggcctaagaaatacttttaag
2814

Query: 7732 tatattttcattagctagaattgcccaatctgtgtaggtataaattacttggtataggga
7791
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2813 tatattttcattagctagaattgcccaatctgtgtaggtataaattacttggtataggga
2754

Query: 7792 gagagaaagcctatcttacctgttgctttcttacttggtggtaacatccagcagttagtc
7851
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2753 gagagaaagcctatcttacctgttgctttcttacttggtggtaacatccagcagttagtc
2694
```

Fig. 5a

```
Query: 7852 tatttataaacataattactttttcacatatgaaccataaaatatttaactttctgctct
7911
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2693 tatttataaacataattactttttcacatatgaaccataaaatatttaactttctgctct
2634

Query: 7912 atattgtttgtttaccgctgtatctcccacagcttgaacagtaccaaggtacgtagtagg
7971
            ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2633 atattgtttgtctaccgctgtatctcccacagcttgaacagtaccaaggtacgtagtagg
2574

Query: 7972 tgctcaataaatgactattgaataaatgaacatatccaacaaatgttctcaatgtaaagg
8031
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2573 tgctcaataaatgactattgaataaatgaacatatccaacaaatgttctcaatgtaaagg
2514

Query: 8032 atcagagatgccacatgttctccttgatgggagagacccttccacatgggaatgatggga
8091
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2513 atcagagatgccacatgttctccttgatgggagagacccttccacatgggaatgatggga
2454

Query: 8092 aggagttgtactcctggatgttcagtaactgcttctaggagaaaaggtagagtcctatca
8151
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2453 aggagttgtactcctggatgttcagtaactgcttctaggagaaaaggtagagtcctatca
2394

Query: 8152 ctaagccgcagatatttatttgtgtgtggctagaatgggatgttttgaatcttctgttac
8211
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2393 ctaagccgcagatatttatttgtgtgtggctagaatgggatgttttgaatcttctgttac
2334

Query: 8212 aaccttgggaacgtggctgttatttcaatttatgagccagaaattttcacatcccgaaac
8271
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2333 aaccttgggaacgtggctgttatttcaatttatgagccagaaattttcacatcccgaaac
2274

Query: 8272 t 8272
            |
Sbjct: 2273 t 2273
```

Score = 1655 bits (835), Expect = 0.0
Identities = 849/856 (99%)
Strand = Plus / Minus

Fig. 5a continued

```
Query: 5903 agatattgctttaggggtatttgatgtggtggtgacggacccctcatgcccagcctcggt
5962
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4642 agatattgctttaggggtatttgatgtggtggtgacggacccctcatgcccagcctcggt
4583


Query: 5963 gctgaagtgtgctgaagcattgcagctgcctgtggtgtcacaagagtgggtgatccagtg
6022
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4582 gctgaagtgtgctgaagcattgcagctgcctgtggtgtcacaagagtgggtgatccagtg
4523


Query: 6023 cctcattgttggggagagaattggattcaagcagcatccaaaatataaacacgattatgt
6082
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4522 cctcattgttggggagagaattggattcaagcagcatccaaaatataaacacgattatgt
4463


Query: 6083 ttctcactaaagatacttggtcttactggttttattccctgctatcgtggagattgtgtt
6142
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4462 ttctcactaaagatacttggtcttactggttttattccctgctatcgtggagattgtgtt
4403


Query: 6143 ttaaccaggttttaaatgtgtcttgtgtgtaactggattccttgcatggatcttgtatat
6202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4402 ttaaccaggttttaaatgtgtcttgtgtgtaactggattccttgcatggatcttgtatat
4343


Query: 6203 agttttatttgctgaactttt atgataaaataaatgttgaatctctttggttgtagtaac
6262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4342 agttttatttgctgaacttttatgataaaataaatgttgaatctctttggttgtagtaac
4283


Query: 6263 tgggatttcttcatctgnnnnnnngagcttaatctcagaacaaatgacaagacatagtac
6322
            |||||||||||||||||       ||||||||||||||||||||||||||||||||||||
Sbjct: 4282 tgggatttcttcatctgtttttttgagcttaatctcagaacaaatgacaagacatagtac
4223


Query: 6323 tttctctgagtctttcaacaggcttattcacttacggaggacagctcaccaaggaaattg
6382
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4222 tttctctgagtctttcaacaggcttattcacttacggaggacagctcaccaaggaaattg
4163


Query: 6383 aaaagttaagagtgaactttattctgtggcatcattcccaaaaggttattccagggtgtc
6442
```

Fig. 5a continued

```
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4162 aaaagttaagagtgaactttattctgtggcatcattcccaaaaggttattccagggtgtc
4103

Query: 6443 taaaatgctatgcttgcagaaactcagtttaaggtaggtgaaggcccagattaacagttg
6502
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4102 taaaatgctatgcttgcagaaactcagtttaaggtaggtgaaggcccagattaacagttg
4043

Query: 6503 tgccaaaagttgagtggaattgggcacagctctgtttcctgacagttaaaaaagacctca
6562
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4042 tgccaaaagttgagtggaattgggcacagctctgtttcctgacagttaaaaaagacctca
3983

Query: 6563 tgctctctctctgagctgagatcacagctcacctgtgggtactccccaactcttagagct
6622
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3982 tgctctctctctgagctgagatcacagctcacctgtgggtactccccaactcttagagct
3923

Query: 6623 aaagggagaacgaaaggaccaactgccatgaagggacagtgaccataagcttgatggaat
6682
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3922 aaagggagaacgaaaggaccaactgccatgaagggacagtgaccataagcttgatggaat
3863

Query: 6683 gaccttccgtaagataaacatgggaagcacaagtgagaacacctggaaatgttacacgtt
6742
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3862 gaccttccgtaagataaacatgggaagcacaagtgagaacacctggaaatgttacacgtt
3803

Query: 6743 ctagtcaaagacccaa 6758
                ||||||||||||||||
Sbjct: 3802 ctagtcaaagacccaa 3787

 Score = 1211 bits (611), Expect = 0.0
 Identities = 625/632 (98%)
 Strand = Plus / Minus

Query: 6778 gtcacaatagctggaagcagttccttcccttcctctggcatcactgatccctgcatggct
6837
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3767 gtcacaatagctggaagcagttccttcccttcctctggcatcactgatccctgcatggct
3708
```

Fig. 5a continued

```
Query: 6838 tctcattctctaaagcaggggtcaacaaggnnnnnnnctgtaaagggtcaaagagtaaat
6897
            ||||||||||||||||||||||||||||||       ||||||||||||||||||||||
Sbjct: 3707 tctcattctctaaagcaggggtcaacaaggtttttttctgtaaagggtcaaagagtaaat
3648

Query: 6898 atttcaggctttgtgggccatttgatccatcacaactactcgcctttgctgtgagggcat
6957
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3647 atttcaggctttgtgggccatttgatccatcacaactactcgcctttgctgtgagggcat
3588

Query: 6958 gaaagcaaccatagacaatgagtaaacaaatgggcacggctgtgtttcagtaaaactgta
7017
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3587 gaaagcaaccatagacaatgagtaaacaaatgggcacggctgtgtttcagtaaaactgta
3528

Query: 7018 caaaaacagacagcaggccatagtttgccagctcctgctccagagacagcagtggaaagg
7077
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3527 caaaaacagacagcaggccatagtttgccagctcctgctccagagacagcagtggaaagg
3468

Query: 7078 gtgatctttagttgataatagcagggaataagttgtcagagcttcccagtgtgtgtagaa
7137
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3467 gtgatctttagttgataatagcagggaataagttgtcagagcttcccagtgtgtgtagaa
3408

Query: 7138 tatgtagtgatgaaaaccagatgcagtgactataacctgatgccagaacactgcattctt
7197
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3407 tatgtagtgatgaaaaccagatgcagtgactataacctgatgccagaacactgcattctt
3348

Query: 7198 tttcagtttggagggcgttgttcagtgaatatttctttttacttacactgatatgaatat
7257
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3347 tttcagtttggagggcgttgttcagtgaatatttctttttacttacactgatatgaatat
3288

Query: 7258 tgattaccagtgatggctgggccatattaagataacttcaacccctatggtttgtgtaag
7317
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3287 tgattaccagtgatggctgggccatattaagataacttcaacccctatggtttgtgtaag
3228

Query: 7318 atgggtaattgggcctgcaatcttcagtatttaaaaatctaacaacttgatctcaatttt
7377
```

Fig. 5a continued

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3227 atgggtaattgggcctgcaatcttcagtatttaaaaatctaacaacttgatctcaatttt
3168

Query: 7378 ttcttaaggaccttttcttggagaataatac 7409
            ||||||||||||||||||||||||||||||||
Sbjct: 3167 ttcttaaggaccttttcttggagaataatac 3136

 Score =  404 bits (204), Expect = e-115
 Identities = 204/204 (100%)
 Strand = Plus / Minus

Query: 5556 cagtgtaacacagcttaccagtgtcttctaattgcggatcagcattgtcgaacccggaag
5615
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6169 cagtgtaacacagcttaccagtgtcttctaattgcggatcagcattgtcgaacccggaag
6110

Query: 5616 tacttcctgtgccttgccagtgggattccttgtgtgtctcatgtctgggtccatgatagt
5675
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6109 tacttcctgtgccttgccagtgggattccttgtgtgtctcatgtctgggtccatgatagt
6050

Query: 5676 tgccatgccaaccagctccagaactaccgtaattatctgttgccagctgggtacagcctt
5735
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6049 tgccatgccaaccagctccagaactaccgtaattatctgttgccagctgggtacagcctt
5990

Query: 5736 gaggagcaaagaattctggactgg 5759
            ||||||||||||||||||||||||
Sbjct: 5989 gaggagcaaagaattctggactgg 5966

 Score =  291 bits (147), Expect = 1e-80
 Identities = 147/147 (100%)
 Strand = Plus / Minus

Query: 5758 ggcaaccccgtgaaaatcctttccagaatctgaaggtactcttggtatcagaccaacagc
5817
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5159 ggcaaccccgtgaaaatcctttccagaatctgaaggtactcttggtatcagaccaacagc
5100

Query: 5818 agaacttcctggagctctggtctgagatcctcatgactggtggtgcagcctctgtgaagc
5877
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5099 agaacttcctggagctctggtctgagatcctcatgactggtggtgcagcctctgtgaagc
5040
```

Fig. 5a continued

```
Query: 5878 agcaccattcaagtgcccataacaaag 5904
            |||||||||||||||||||||||||||
Sbjct: 5039 agcaccattcaagtgcccataacaaag 5013


 Score =  281 bits (142), Expect = 9e-78
 Identities = 142/142 (100%)
 Strand = Plus / Minus


Query: 8920 ctgcccagagttccaccagcctgggtatagtatttgttataatctagtcgtaacagtagt
8979
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2274 ctgcccagagttccaccagcctgggtatagtatttgttataatctagtcgtaacagtagt
2215


Query: 8980 tgagccaaatctgagttgatctgatgattccgaacactggagagaatcttgaacaggagt
9039
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2214 tgagccaaatctgagttgatctgatgattccgaacactggagagaatcttgaacaggagt
2155


Query: 9040 gaagactggcggctaaagccct 9061
            ||||||||||||||||||||||
Sbjct: 2154 gaagactggcggctaaagccct 2133


 Score =  226 bits (114), Expect = 5e-61
 Identities = 117/118 (99%)
 Strand = Plus / Minus


Query: 9673 ccttcacgagaatgctcagctgggcggctccacgctcatccagtgggcctaggttctgac
9732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2135 ccttcacgagaatgctcagctgggcggctccacgctcatccagtgggcctaggttctgac
2076


Query: 9733 tgaccagcgaacaaaaactgtgacagagatctaggatttcattcaggcagtgaaacac 9790
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2075 tgaccagcaaacaaaaactgtgacagagatctaggatttcattcaggcagtgaaacac 2018


 Score =  190 bits (96), Expect = 3e-50
 Identities = 96/96 (100%)
 Strand = Plus / Minus


Query: 5463 gaatttttggaaattcctcctttcaacaagcagtatacagaatcccagcttcgagcagga
5522
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6812 gaatttttggaaattcctcctttcaacaagcagtatacagaatcccagcttcgagcagga
6753
```

Fig. 5a continued

```
Query: 5523 gctggctatatccttgaagatttcaatgaagcccag 5558
            ||||||||||||||||||||||||||||||||||||
Sbjct: 6752 gctggctatatccttgaagatttcaatgaagcccag 6717

 Score = 52.0 bits (26), Expect = 2e-08
 Identities = 26/26 (100%)
 Strand = Plus / Minus

Query: 7668 gcctcccaaagtgctgggattacagg 7693
            ||||||||||||||||||||||||||
Sbjct: 5312 gcctcccaaagtgctgggattacagg 5287
```

Fig. 5a continued

```
CIDEB1_BLTR2      CIDEB1      2289  BLTR2 6530  OL:   2254  OF1:  17    OF2:
   1

 Score = 2727 (753.8 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 547/549 (99%), Positives = 547/549 (99%),  Strands Minus /
Plus

Query:  2250 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT
2191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT 60

Query:  2190 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
2131
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
120

Query:  2130 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
2071
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
180

Query:  2070 TGGGAGGCTAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
2011
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   181 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
240

Query:  2010 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
1951
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   241 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
300

Query:  1950 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
1891
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   301 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
360

Query:  1890 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
1831
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
420

Query:  1830 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
1771
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   421 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
480

Query:  1770 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC
1711
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   481 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC
540
```

Fig. 5b

```
Query:  1710 TCCCTGAGT 1702
             ||||||| |
Sbjct:   541 TCCCTGAAT 549

 Score = 1322 (365.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 266/268 (99%), Positives = 266/268 (99%), Strands Minus /
Plus

Query:   757 CCTGTAGGCCCAGAAGGATGTCGGTCTGCTACCGTCCCCCAGGGAACGAGACACTGCTGA
698
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5426 CCTGTAGGCCCAGAAGGATGTCGGTCTGCTACCGTCCCCCAGGGAACGAGACACTGCTGA
5485

Query:   697 GCTGGAAGACTTCGCGGGCCACAGGCACAGCCTTCCTGCTGCTGGCGGCGCTGCTGGGGC
638
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5486 GCTGGAAGACTTCGCGGGCCACAGGCACAGCCTTCCTGCTGCTGGCGGCGCTGCTGGGGC
5545

Query:   637 TGCCTCCCAACCCCTTCGTGGTGTGGAGCTTGCCGGGCTGGCAGCCTCCACCCCGGCCAC
578
             ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct:  5546 TGCCTGGCAACGGCTTCGTGGTGTGGAGCTTGGCGGGCTGGCGGCCTGCACGGGGGCGAC
5605

Query:   577 CGCTGGCGGCCACGCTTGTGCTGCACCTGGCGCTGGCCGACGGCGCGGTGCTGCTGCTCA
518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5606 CGCTGGCGGCCACGCTTGTGCTGCACCTGGCGCTGGCCGACGGCGCGGTGCTGCTGCTCA
5665

Query:   517 CGCCGCTCTTTGTGGCCTTCCTGACCGG 490
             ||||||||||||||||||||||||| |
Sbjct:  5666 CGCCGCTCTTTGTGGCCTTCCTGACCCG 5693

 Score = 1316 (363.8 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 264/265 (99%), Positives = 264/265 (99%), Strands Minus /
Plus

Query:   421 CAAGCGTGCTGCTCACCGGCCTCCTCACCCTGCAGCGCTGCCTCGCAGTCACCCGCCCCT
362
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5762 CCAGCGTGCTGCTCACCGGCCTGCTCAGCCTGCAGCGCTGCCTCGCAGTCACCCGCCCCT
5821

Query:   361 TCCTGGCGCCTCGGCTGCGCAGCCCGGCCCTGGCCCGCCGCCTGCTGCTGGCGGTCTGGC
302
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5822 TCCTGGCGCCTCGGCTGCGCAGCCCGGCCCTGGCCCGCCGCCTGCTGCTGGCGGTCTGGC
5881

Query:   301 TGGCCGCCCTGTTGCTCGCCGTCCCGGCCGCCGTCTACCGCCACCTGTGGAGGGACCGCG
242
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5862 TGGCCGCCCTGTTGCTCGCCGTCCCGGCCGCCGTCTACCGCCACCTGTGGAGGGACCGCG
5941
```

Fig. 5b continued

```
Query:   241 TATGCCAGCTGTGCCACCCGTCGCCGGTCCACGCCGCCGCCCACCTGAGCCTGGAGACTC
182
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5942 TATGCCAGCTGTGCCACCCGTCGCCGGTCCACGCCGCCGCCCACCTGAGCCTGGAGACTC
6001

Query:   181 TGACCGCTTTCGTGCTTCCTTTCGG 157
             |||||||||||||||||||||||||
Sbjct:  6002 TGACCGCTTTCGTGCTTCCTTTCGG 6026

 Score = 920 (254.3 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 188/193 (97%), Positives = 188/193 (97%),  Strands Minus /
Plus

Query:  1708 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
1649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   724 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783

Query:  1648 GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
1589
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   784 GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
843

Query:  1588 AAAGGTGAATCGGCCGATGTCCTTGCTGTGCTTGGGGCTCTCCCGTCCAGGCCCATATGA
1529
             ||||||||||||| |||||||||||||||||||||| |||||||||||  | ||||||||
Sbjct:   844 AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCATATGA
903

Query:  1528 CAGCACTCCACTC 1516
             |||||||||||||
Sbjct:   904 CAGCACTCCACTC 916

 Score = 753 (208.2 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 157/165 (95%), Positives = 157/165 (95%),  Strands Minus /
Plus

Query:  1529 ACAGCACTCCACTCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCATCAGGCACG
1470
             | ||  |||    |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1139 ATAGGCCTCTTACCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCATCAGGCACG
1198

Query:  1469 TGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATCCTCCTCTA
1410
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1199 TGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATCCTCCTCTA
1258

Query:  1409 GCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCTT 1365
             ||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  1259 GCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCCT 1303

 Score = 746 (206.2 bits), Expect = 0.0, Sum P(13) = 0.0
```

Fig. 5b continued

```
 Identities = 150/151 (99%), Positives = 150/151 (99%), Strands Minus /
Plus

Query:  1369 TGCTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
1310
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2146 TACTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
2205

Query:  1309 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
1250
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2206 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
2265

Query:  1249 AAACTCCGAGCTTATATTAGATACTGACCTG 1219
             |||||||||||||||||||||||||||||||
Sbjct:  2266 AAACTCCGAGCTTATATTAGATACTGACCTG 2296

 Score = 737 (203.7 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 148/150 (98%), Positives = 149/150 (99%), Strands Minus /
Plus

Query:  1118 CTGCTCTTTCCTTCCTCCTTGGTCGGAGGAGGGGCTGGCTCACTGCTCTGGCTTCATTTT
1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3257 CTGCTCTTTCCTTCCTCCTTGGTCGGAGGAGGGGCTGGCTCACTGCTCTGGCTTCATTTT
3316

Query:  1058 CCAGAGCTGCCTGCTGCAGTCACACTTAGGTCATCTTCTCTCACTTTTCTCCTTTTGCCG
999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3317 CCAGAGCTGCCTGCTGCAGTCACACTTAGGTCATCTTCTCTCACTTTTCTCCTTTTGCCG
3376

Query:   998 ATTAGTGGACGTGACAGAGATGTGAATRTG 969
             ||||||||||||||||||||||||||+ |
Sbjct:  3377 ATTAGTGGACGTGACAGAGATGTGAATGGG 3406

 Score = 714 (197.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 146/150 (97%), Positives = 146/150 (97%), Strands Minus /
Plus

Query:   166 TTCCTTTCGGCTGATGCTCGGCTGCTACAGCGTGACGCTGGCACGGCTGCGGGGCGCCCG
107
             | | ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  6018 TCCTTTCGGGCTGATGCTCGGCTGCTACAGCGTGACGCTGGCACGGCTGCGGGGCGCCCG
6077

Query:   106 CTGGGGCTCCGGGCGGCACGGGGCGCGGGTGGGCCGGCTGGTGAGCGCCATCGTGCTTGC 47
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  6078 CTGGGGCTCCGGGCGGCACGGGGCGCGGGTGGGCCGGCTGGTGAGCGCCATCGTGCTTGC
6137

Query:    46 CTTCGGCTTGCTCTGGGCCCCCTACCACGC 17
             ||||||||||||||||||||||||||||||
Sbjct:  6138 CTTCGGCTTGCTCTGGGCCCCCTACCACGC 6167
```

Fig. 5b continued

```
 Score = 638 (176.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 130/133 (97%), Positives = 130/133 (97%),  Strands Minus /
Plus

Query:   962 GGGCAGGGGATGTCCTTTGATGGCATCAAGACTTTAGCTTCTGGTGCGCTGTGTCCCAGC
903
             |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3404 GGGGCAGGGATGTCCTTTGATGGCATCAAGACTTTAGCTTCTGGTGCGCTGTGTCCCAGC
3463

Query:   902 TCTGATTTCAGTTGCAGCCGTGATGGACAGTTGCATGGAAGCTGAGACTCTCACTGACAG
843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3464 TCTGATTTCAGTTGCAGCCGTGATGGACAGTTGCATGGAAGCTGAGACTCTCACTGACAG
3523

Query:   842 TGAAACCCTCAAA 830
             |||||||||||||
Sbjct:  3524 TGAAACCCTCAAA 3536

 Score = 537 (148.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 109/111 (98%), Positives = 109/111 (98%),  Strands Minus /
Plus

Query:  1227 ACTGACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
1168
             | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2590 AGTCACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
2649

Query:  1167 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCT 1117
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2650 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCT 2700

 Score = 394 (108.9 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 82/86 (95%), Positives = 82/86 (95%),  Strands Minus / Plus

Query:   839 AACCCTCAAAATGAACACAATCCCTGCTTTCCTGCCAAGGATCCTTGTAGGGTCCCCCAG
780
             || ||    |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3526 AAACCCTCAAATGAACACAATCCCTGCTTTCCTGCCAAGGATCCTTGTAGGGTCCCCCAG
3585

Query:   779 CTTCCCCACTTTTTTTCTGTGTCCTG 754
             ||||||||||||||||||||||||||
Sbjct:  3586 CTTCCCCACTTTTTTTCTGTGTCCTG 3611

 Score = 370 (102.3 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 74/74 (100%), Positives = 74/74 (100%),  Strands Minus / Plus

Query:   493 CCGGCAGGCCTGGCCGCTGGGCCAGGCGGGCTGCAAGGCGGTGTACTACGTGTGCGCGCT
434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5691 CCGGCAGGCCTGGCCGCTGGGCCAGGCGGGCTGCAAGGCGGTGTACTACGTGTGCGCGCT
5750

Query:   433 CAGCATGTACGCCA 420
             ||||||||||||||
Sbjct:  5751 CAGCATGTACGCCA 5764
```

Fig. 5b continued

CIDEB2_BLTR2 CIDEB2 1511 BLTR2 6530 OL: 1410 OF1: 1 OF2: 1

Score = 2736 (756.0 bits), Expect = 0.0, Sum P(5) = 0.0
Identities = 548/549 (99%), Positives = 548/549 (99%), Strands Minus / Plus

```
Query: 1384 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT 1325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    1 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT 60

Query: 1324 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA 1265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   61 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA 120

Query: 1264 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG 1205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  121 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG 180

Query: 1204 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG 1145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  181 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG 240

Query: 1144 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA 1085
          : ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  241 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA 300

Query: 1084 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC 1025
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  301 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC 360

Query: 1024 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA 965
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  361 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA 420

Query:  964 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG 905
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  421 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG 480

Query:  904 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC 845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  481 GAAATTCCCAGCAACATATGGCCCA                           ACGAAGGAGC 540
```

Fig. 5c

```
Query:    844 TCCCTGAGT 836
              ||||||| |
Sbjct:    541 TCCCTGAAT 549

 Score = 1787 (493.8 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 359/361 (99%), Positives = 359/361 (99%),  Strands Minus /
Plus

Query:    361 ACTGACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
302
              | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2590 AGTCACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
2649

Query:    301 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCTTTAGTGGTG
242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2650 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCTTTAGTGGTG
2709

Query:    241 TTTTCTGTTACAAACCTGGGATCTCAGCCCAGGACAAGGTGGGAATGAGTCAAGCCTGGA
182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2710 TTTTCTGTTACAAACCTGGGATCTCAGCCCAGGACAAGGTGGGAATGAGTCAAGCCTGGA
2769

Query:    181 CTCTGGCCCCCCTGCCTGGCCAGTAAGAAGGGCAAAGTCCAAGGGGAGGGATGAGGGAGG
122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2770 CTCTGGCCCCCCTGCCTGGCCAGTAAGAAGGGCAAAGTCCAAGGGGAGGGATGAGGGAGG
2829

Query:    121 GGCCAGATGGGGTCCTGGAGGAAGAATTGCCTGGCAAAAGCCATTGGAGCTTGTATGTGT  62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2830 GGCCAGATGGGGTCCTGGAGGAAGAATTGCCTGGCAAAAGCCATTGGAGCTTGTATGTGT
2889

Query:     61 GTCTTTGGTGATGACATGTGTTGTGAGGGTAGATGGGAACCATGTAAAAGGATGAAATGT  2
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2890 GTCTTTGGTGATGACATGTGTTGTGAGGGTAGATGGGAACCATGTAAAAGGATGAAATGT
2949

Query:      1 G 1
              |
Sbjct:   2950 G 2950

 Score = 965 (266.6 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 193/193 (100%), Positives = 193/193 (100%),  Strands Minus /
Plus

Query:    842 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    724 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783

Query:    782 GAATCTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
723
```

Fig. 5c continued

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    784 GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
843

Query:    722 AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCATATGA
663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    844 AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCATATGA
903

Query:    662 CAGCACTCCACTC 650
              |||||||||||||
Sbjct:    904 CAGCACTCCACTC 916

 Score = 757 (209.2 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 161/173 (93%), Positives = 161/173 (93%),  Strands Minus /
Plus

Query:    671 GCCATATGACAGCACTCCACTCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCAT
612
              |||  |  | ||  |||      ||||||||||||||||||||||||||||||||||||||
Sbjct:   1131 GCCCCAGTATAGGCCTCTTACCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCAT
1190

Query:    611 CAGGCACGTGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATC
552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1191 CAGGCACGTGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATC
1250

Query:    551 CTCCTCTAGCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCTT 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:   1251 CTCCTCTAGCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCCT 1303

 Score = 746 (206.1 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 150/151 (99%), Positives = 150/151 (99%),  Strands Minus /
Plus

Query:    503 TGCTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
444
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2146 TACTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
2205

Query:    443 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2206 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
2265

Query:    383 AAACTCCGAGCTTATATTAGATACTGACCTG 353
              |||||||||||||||||||||||||||||||
Sbjct:   2266 AAACTCCGAGCTTATATTAGATACTGACCTG 2296
```

Fig. 5c continued

```
APAF1_EB1    APAF1    7042    EB1a    1752    OL: 141   OF1: 6889
OF2: 1612

 Score = 705 (194.8 bits), Expect = 3.9e-52, P = 3.9e-52
 Identities = 141/141 (100%), Positives = 141/141 (100%),  Strands Minus /
Plus

Query:  7029 TGTTTTTCAAAACAATTTTGTGAATTTTATTTTTACAAAAATTTTTTAAATTCATATTTT
6970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1612 TGTTTTTCAAAACAATTTTGTGAATTTTATTTTTACAAAAATTTTTTAAATTCATATTTT
1671

Query:  6969 AAAATGTATACCAAGGCAAAAAAATCATATAAGCTATATCATAAATACAAGAGTTTCAAA
6910
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1672 AAAATGTATACCAAGGCAAAAAAATCATATAAGCTATATCATAAATACAAGAGTTTCAAA
1731

Query:  6909 ACATACAAGAGACATATAATG 6889
             |||||||||||||||||||||
Sbjct:  1732 ACATACAAGAGACATATAATG 1752
```

Fig. 5d

```
Hum_AChR_MINK2    AChR    2457    MINK2    4863    OL: 236 OF1: 2175
OF2: 4583

 Score =  218 bits (110), Expect = 2e-59
 Identities = 110/110 (100%)
 Strand = Plus / Minus


Query: 2254 aagggttacttgctgctcacactatatacagatgcaagcaaggggcgtggagagtgaggg
2313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4787 aagggttacttgctgctcacactatatacagatgcaagcaaggggcgtggagagtgaggg
4728


Query: 2314 ctccctgctccctccctccaccggggaagggcatgggctagaagaggaga 2363
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4727 ctccctgctccctccctccaccggggaagggcatgggctagaagaggaga 4678


 Score =  133 bits (67), Expect = 9e-34
 Identities = 74/75 (98%), Gaps = 1/75 (1%)
 Strand = Plus / Minus


Query: 2384 aatgttttggctg-cggggtccccccctccattccctggagtttggggaaggggaatcat
2442
            ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4657 aatgttttggctggcggggtccccccctccattccctggagtttggggaaggggaatcat
4598


Query: 2443 taaagtgctttcaga 2457
            |||||||||||||||
Sbjct: 4597 taaagtgctttcaga 4583


 Score =  103 bits (52), Expect = 8e-25
 Identities = 52/52 (100%)
 Strand = Plus / Minus


Query: 2175 agctggttgaattgtctttattaacaaacaggatatccaaggccactacatt 2226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4863 agctggttgaattgtctttattaacaaacaggatatccaaggccactacatt 4812
```

Fig. 5e

```
Mus_AChR    Mus_AChR     1590    Anti_Mus_AChR         2227      OL: 672   OF1:
934         OF2: 506

 Score = 1221 (337.4 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 245/246 (99%), Positives = 245/246 (99%),  Strands Minus /
Plus

Query:   1587 AGTTCACAAACCAGATTTATTGTCAGCGGCCTGTTTTCAAAATCTCTTTCTTGGGGGGTG
1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    506 AGTTCACAAACCAGATTTATTGTCAGCGGCCTGTTTTCAAAATCTCTTTCTTGGGGGGTG
565

Query:   1527 GGGGAGAGGTGGGTGCCAGTGCAGGCTCATGGTTGGATGCACGGTGGGTAAGGGAGATCA
1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    566 GGGGAGAGGTGGGTGCCAGTGCAGGCTCATGGTTGGATGCACGGTGGGTAAGGGAGATCA
625

Query:   1467 GGAACTTGGTTGAAGTAACCCCCAAGGAAGATGAGAGTAGAACCAACGCTGAAGAGCACC
1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    626 GGAACTTGGTTGAAGTAACCCCCAAGGAAGATGAGAGTAGAACCAACGCTGAAGAGCACC
685

Query:   1407 AAAGCTGCCCAAAAACAGACATTGTCCAGGGCCTTCCCCATACGCACCCAGTCGGACAGT
1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    686 AAAGCTGCCCAAAAACAGACATTGTCCAGGGCCTTCCCCATACGCACCCAGTCGGACAGT
745

Query:   1347 TCCTCT 1342
              |||| |
Sbjct:    746 TCCTGT 751

 Score = 954 (263.6 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 198/207 (95%), Positives = 198/207 (95%),  Strands Minus /
Plus

Query:   1249 GCAGAGGGCTGCGGTCCAAGTTCCGTGCCGATGCCTCTGACCCTCAAACACGAGTTCGCT
1190
              |||| |      ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1007 GCAGTGCTTACCGGTCCAAGTTCCGTGCCGATGCCTCTGACCCTCAAACACGAGTTCGCT
1066

Query:   1189 CCGCGGCTTTTTCAAGATGAGCTCCTCCGCTCTGAGCAGAATGCCCACAGACGAGGCACG
1130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1067 CCGCGGCTTTTTCAAGATGAGCTCCTCCGCTCTGAGCAGAATGCCCACAGACGAGGCACG
1126

Query:   1129 CCTCGCTGGTGAGGCAGTTCGGGGATCCTCTGGGGGTGGGCTCGAGCCCAGGAGACGCGG
1070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1127 CCTCGCTGGTGAGGCAGTTCGGGGATCCTCTGGGGGTGGGCTCGAGCCCAGGAGACGCGG
1186

Query:   1069 CAGCAGCTCTAATAAAATCTGGCGCAG 1043
              |||||||||||||||||||||   || |
```

Fig. 5f

```
Sbjct:  1187 CAGCAGCTCTAATAAAATCTGCAGCCG 1213

 Score = 591 (163.3 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 119/120 (99%), Positives = 119/120 (99%),  Strands Minus /
Plus

Query:  1053 ATCTGGCGCAGCCGAGGGGATGTAGCATGAGTCGTTGGCGTCCTCAAAGATACGTTGAGC
994
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1277 ACCTGGCGCAGCCGAGGGGATGTAGCATGAGTCGTTGGCGTCCTCAAAGATACGTTGAGC
1336

Query:   993 ACGATGACGCAATTCATGACAATGAGCGTGGCAACCACCATGACGAATATAAGATACCTG
934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1337 ACGATGACGCAATTCATGACAATGAGCGTGGCAACCACCATGACGAATATAAGATACCTG
1396

 Score = 546 (150.9 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 110/111 (99%), Positives = 110/111 (99%),  Strands Minus /
Plus

Query:  1346 CCTCTCCAGTGGCTTCCTGGTCTCTTGTGCTCTCAGCCACAAAGTTCACAGCATCCACAC
1287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   826 CCTCTCCAGTGGCTTCCTGGTCTCTTGTGCTCTCAGCCACAAAGTTCACAGCATCCACAC
885

Query:  1286 AGCAGCGGATTTCTGGGGCTGCAGCACCCAGGTTCTGGCAGAGGGCTGCGG 1236
             ||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:   886 AGCAGCGGATTTCTGGGGCTGCAGCACCCAGGTTCTGGCAGAGGGCTGCTG 936
```

Fig. 5f continued

```
CyclinE2   CyclinE2    2714    Anti_CyclinE2         6773      OL: 1855
OF1: 865         OF2: 2006

 Score = 7885 (2178.8 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 1577/1577 (100%), Positives = 1577/1577 (100%),  Strands Minus
/ Plus

Query:  2714 TACAGCTGGCAGCGCAGAGAAGGAAAAAAAAGTTTCTCCAAGCAATGGCAAAACTTTACT
2655
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2006 TACAGCTGGCAGCGCAGAGAAGGAAAAAAAAGTTTCTCCAAGCAATGGCAAAACTTTACT
2065

Query:  2654 TTTAAGCAGTTAAATTTTTTAACTTTTATTTTTAAACAATGGGCTAAAAATAAACAGT
2595
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2066 TTTAAGCAGTTAAATTTTTTAACTTTTATTTTTAAACAATGGGCTAAAAATAAACAGT
2125

Query:  2594 ATTAAAAGGTTAAGTTTATATAATACATATGTACACAATTAGTGGTGTTTTCTTTTCAGA
2535
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2126 ATTAAAAGGTTAAGTTTATATAATACATATGTACACAATTAGTGGTGTTTTCTTTTCAGA
2185

Query:  2534 CAAAATACTGAAACAAATATTAGTTTAAAAACAAACTATACAGAAGACTTCATACCGTAA
2475
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2186 CAAAATACTGAAACAAATATTAGTTTAAAAACAAACTATACAGAAGACTTCATACCGTAA
2245

Query:  2474 CAATAAATGTATAGTTTCTTCAAAGGGAGAAGAGATTCACATATCTGATAACAAAATAAA
2415
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2246 CAATAAATGTATAGTTTCTTCAAAGGGAGAAGAGATTCACATATCTGATAACAAAATAAA
2305

Query:  2414 CTAGCAATCTAGTTTTCTAATCTACTTTATGAGGCTGGATTTTTTTTTTAGAAAAGCTAA
2355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2306 CTAGCAATCTAGTTTTCTAATCTACTTTATGAGGCTGGATTTTTTTTTTAGAAAAGCTAA
2365

Query:  2354 TTTAAAATATTTAGAAATAGCTAGCCTATGTACAGCAAGTTTTCATGTCTTTTTTTAATA
2295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2366 TTTAAAATATTTAGAAATAGCTAGCCTATGTACAGCAAGTTTTCATGTCTTTTTTTAATA
2425

Query:  2294 AATAGATTTCTAGGAGTCAGTATATATTTAATACTCTTCTTCCTTAAGAAAATAGAAGTT
2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2426 AATAGATTTCTAGGAGTCAGTATATATTTAATACTCTTCTTCCTTAAGAAAATAGAAGTT
2485

Query:  2234 TAGGTCAAGTGTTAAGCTTTATCACTTTGACACTGTCCTTATCTCACAATGGAGGAATTT
2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Fig. 5g

Sbjct: 2486 TAGGTCAAGTGTTAAGCTTTATCACTTTGACACTGTCCTTATCTCACAATGGAGGAATTT 2545

Query: 2174 AGAAAGGACCTTAACAGTTTCACAAACATAAATAAAGCCTTAGTCACACTAAATTAAAAA 2115
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2546 AGAAAGGACCTTAACAGTTTCACAAACATAAATAAAGCCTTAGTCACACTAAATTAAAAA 2605

Query: 2114 AAAAAATTCCTTAGGGATATCTTAGAGTAGTAAAGTGACTTCCTCATATAAATAGTTTGA 2055
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2606 AAAAAATTCCTTAGGGATATCTTAGAGTAGTAAAGTGACTTCCTCATATAAATAGTTTGA 2665

Query: 2054 AAGGGTACTTAAGTTTTTCACCCAAATTGTGATATACAAAAAGGTTATTACCAAGCAACC 1995
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2666 AAGGGTACTTAAGTTTTTCACCCAAATTGTGATATACAAAAAGGTTATTACCAAGCAACC 2725

Query: 1994 TACATGTCAAGAAAGCCCCAGTTAGGAAGGAGCCACAGCATTTATCTTGTTTATAATTTC 1935
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2726 TACATGTCAAGAAAGCCCCAGTTAGGAAGGAGCCACAGCATTTATCTTGTTTATAATTTC 2785

Query: 1934 TTTGGTACTCCCACTGTTTAGAGCACAGGTTGAACACCATGTTCATCTAAGCCTTATTAG 1875
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2786 TTTGGTACTCCCACTGTTTAGAGCACAGGTTGAACACCATGTTCATCTAAGCCTTATTAG 2845

Query: 1874 TTAAAAAATGTGTTATGGCAAGGCAAATAAACTAGTTTAAAAAACATTAAATTTCACCAT 1815
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2846 TTAAAAAATGTGTTATGGCAAGGCAAATAAACTAGTTTAAAAAACATTAAATTTCACCAT 2905

Query: 1814 TTGTAGAAATTCAAGTTTTATAATAGCTTGCTATAGCAGCTATAGATAAATTAGTCACCT 1755
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2906 TTGTAGAAATTCAAGTTTTATAATAGCTTGCTATAGCAGCTATAGATAAATTAGTCACCT 2965

Query: 1754 TATTACAAAACTAAACCTTTGTAAACAAGTTTAAATTTAATTTTCAAGAACCAAATTGCA 1695
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2966 TATTACAAAACTAAACCTTTGTAAACAAGTTTAAATTTAATTTTCAAGAACCAAATTGCA 3025

Query: 1694 CTAGTCAAGAGTGTAGGAATTTTGAGAATCTAACAACTAGATTCAAAGTACTGTATCACT 1635
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3026 CTAGTCAAGAGTGTAGGAATTTTGAGAATCTAACAACTAGATTCAAAGTACTGTATCACT 3085

Query: 1634 TAGTATACCCTTTAAGGTAGCACTTATCCAGTCCAAAACTCCAGTGACAAAATTCCTAGT 1575

Fig. 5g continued

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3086 TAGTATACCCTTTAAGGTAGCACTTATCCAGTCCAAAACTCCAGTGACAAAATTCCTAGT
3145

Query:    1574 TTATCAAGATAAACACAGTAACACTGGATTAAAGGAAAAACATTGCTATGGTATAGACTG
1515
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3146 TTATCAAGATAAACACAGTAACACTGGATTAAAGGAAAAACATTGCTATGGTATAGACTG
3205

Query:    1514 TGGTTGGCTTCTATCCAGTAACCTTGGGAATGAAGACATCTTTGTAAACAAGTCCTGCTG
1455
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3206 TGGTTGGCTTCTATCCAGTAACCTTGGGAATGAAGACATCTTTGTAAACAAGTCCTGCTG
3265

Query:    1454 TTTCTTTAACAGCTAACATAGGAAATAATTAAATGTATTCTTTAGTGCCAATTGTAAGTT
1395
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3266 TTTCTTTAACAGCTAACATAGGAAATAATTAAATGTATTCTTTAGTGCCAATTGTAAGTT
3325

Query:    1394 TTAAAATCAGAATGGCAGTGTAACTTGTGAATTGGCTAGGGCAATCAATCACAGCACTAC
1335
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3326 TTAAAATCAGAATGGCAGTGTAACTTGTGAATTGGCTAGGGCAATCAATCACAGCACTAC
3385

Query:    1334 TTTCTGTAAAACTTTAGTAGTTCAGTGATACCAGTTCTACCCAATCTTGGTGAATTCCAA
1275
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3386 TTTCTGTAAAACTTTAGTAGTTCAGTGATACCAGTTCTACCCAATCTTGGTGAATTCCAA
3445

Query:    1274 CTTGTTTGCTTAGTTATCTTCTTTAGTGTTTTCCTGGTGGTTTTTCAGTGCTCTTCGGTG
1215
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3446 CTTGTTTGCTTAGTTATCTTCTTTAGTGTTTTCCTGGTGGTTTTTCAGTGCTCTTCGGTG
3505

Query:    1214 GTGTCATAATGCCTCCATTGCACACTGGTGACAACTGTCCCCCTTTTCTGAAGGTGTTTA
1155
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3506 GTGTCATAATGCCTCCATTGCACACTGGTGACAACTGTCCCCCTTTTCTGAAGGTGTTTA
3565

Query:    1154 TGTAATTTACTTCCTCC 1138
                |||||||||||||||||
Sbjct:    3566 TGTAATTTACTTCCTCC 3582

 Score = 805 (222.4 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 161/161 (100%), Positives = 161/161 (100%),  Strands Minus /
Plus

Query:    1139 CCAGCATAGCCAAATAGTTTGTATGTGTCTGGATATTATGTCTGTCTTCCATAGGAATCT
1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Fig. 5g continued

```
Sbjct:   3967 CCAGCATAGCCAAATAGTTTGTATGTGTCTGGATATTATGTCTGTCTTCCATAGGAATCT
4026

Query:   1079 TCTTAAAAGTCTTCAGCTTCACTGCACTAGTACTTTTTACTACATTCACAAAAGCTACCA
1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   4027 TCTTAAAAGTCTTCAGCTTCACTGGACTAGTACTTTTTACTACATTGACAAAAGGTACCA
4086

Query:   1019 TCCAATCTACACATTCTGAAATACTGTCCCACTCCAAACCT 979
              |||||||||||||||||||||||||||||||||||||||||
Sbjct:   4087 TCCAATCTACACATTCTGAAATACTGTCCCACTCCAAACCT 4127

 Score = 581 (160.5 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 117/118 (99%), Positives = 117/118 (99%),  Strands Minus /
Plus

Query:    982 ACCTGAGGCTTTCTTAACCACTTCAATGGAGGTAAAATGGCACAAGGCAGCAGCAGTCAG
923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   4615 ACCTGAGGCTTTCTTAACCACTTCAATGGAGGTAAAATGGCACAAGGCAGCAGCAGTCAG
4674

Query:    922 TATTCTGTACTGGAACTCTAATGAATCAATGGCTAGAATACACAGATCTAAAAGCTGA 865
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:   4675 TATTCTGTACTGGAACTCTAATGAATCAATGGCTAGAATACACAGATCTAAAAGCTAA 4732
```

Fig. 5g continued

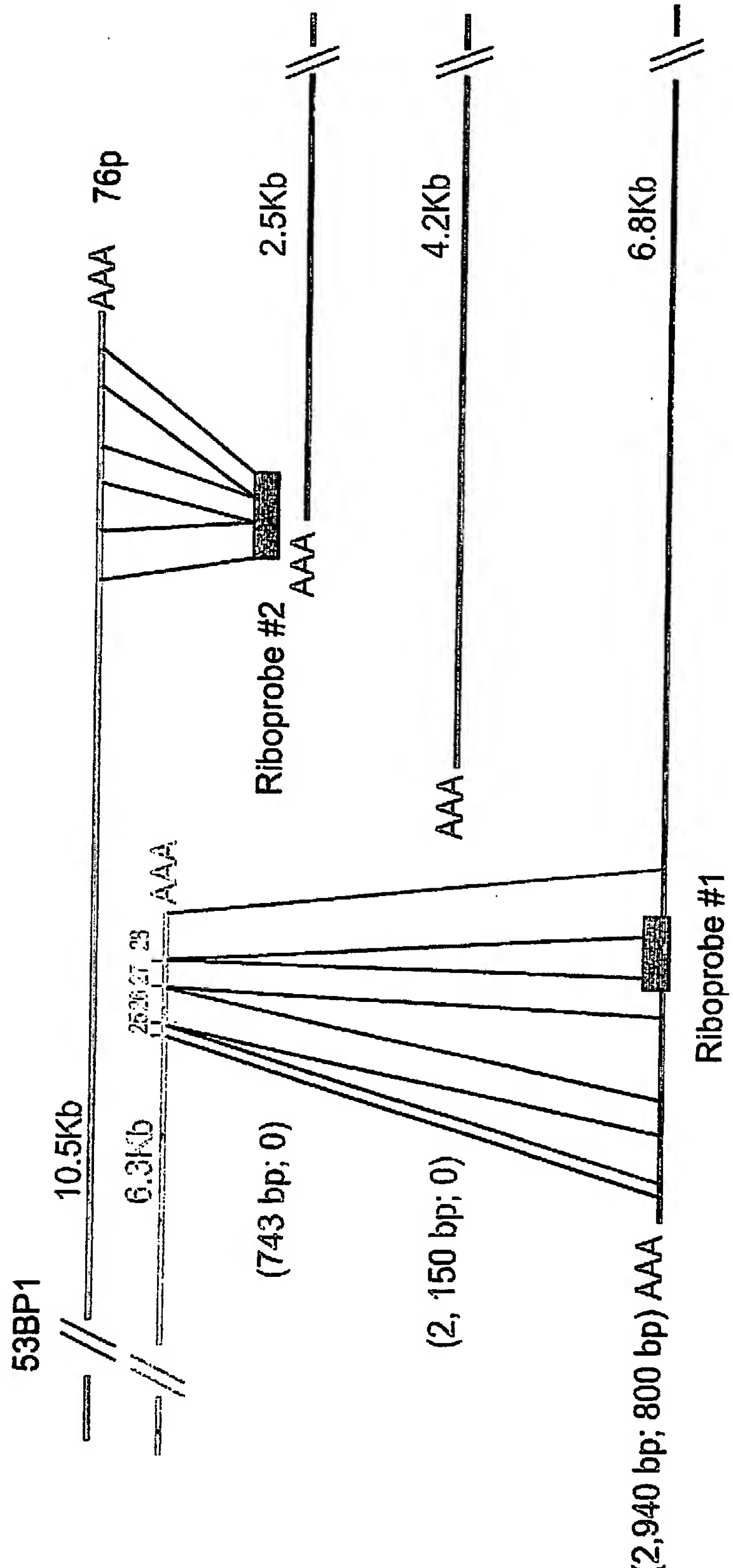

53BP1
10.5Kb
6.3Kb
AAA
76p
2.5Kb
AAA
Riboprobe #2
4.2Kb
AAA
6.8Kb
Riboprobe #1
(743 bp; 0)
(2, 150 bp; 0)
(2,940 bp; 800 bp) AAA

Fig. 6

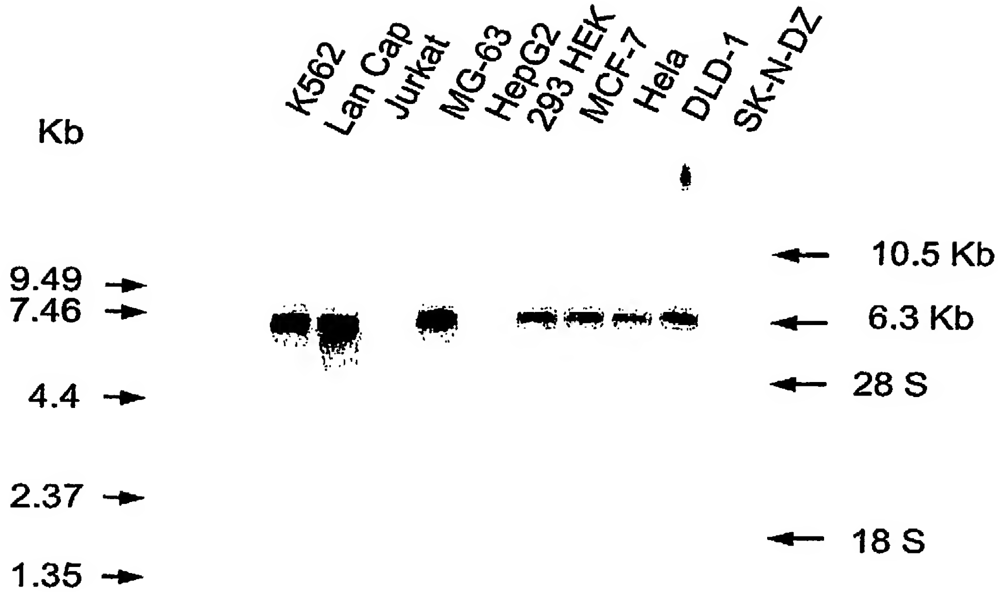

K562
Lan Cap
Jurkat
MG-63
HepG2
293 HEK
MCF-7
Hela
DLD-1
SK-N-DZ
Kb
9.49
7.46
4.4
2.37
1.35
10.5 Kb
6.3 Kb
28 S
18 S

Fig. 7

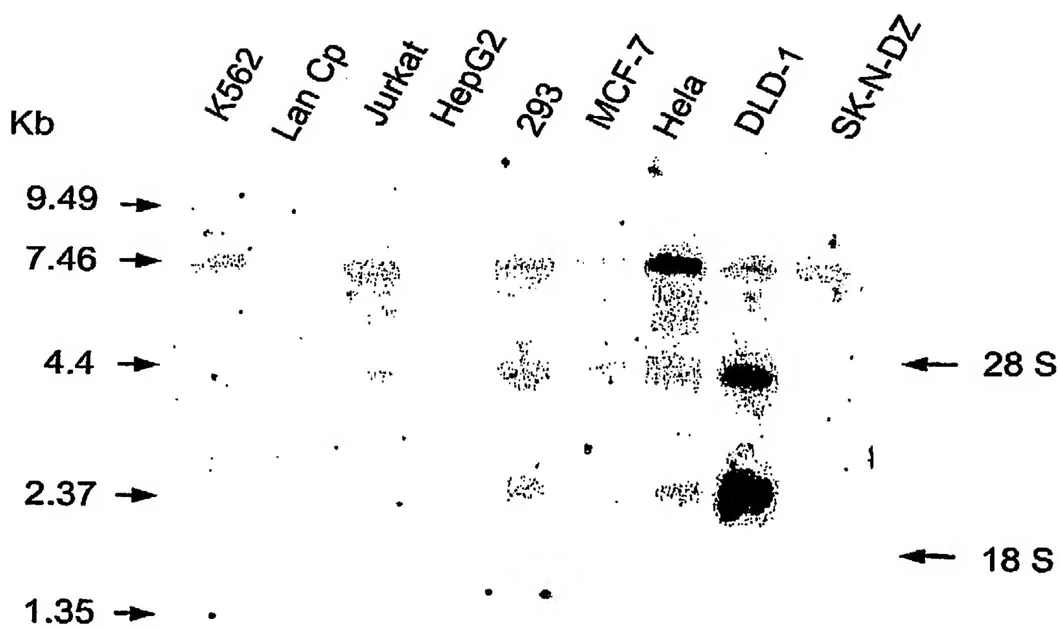

K562
Lan Cp
Jurkat
HepG2
293
MCF-7
Hela
DLD-1
SK-N-DZ
Kb
9.49
7.46
4.4
2.37
1.35
28 S
18 S

Fig. 8

Fig. 9

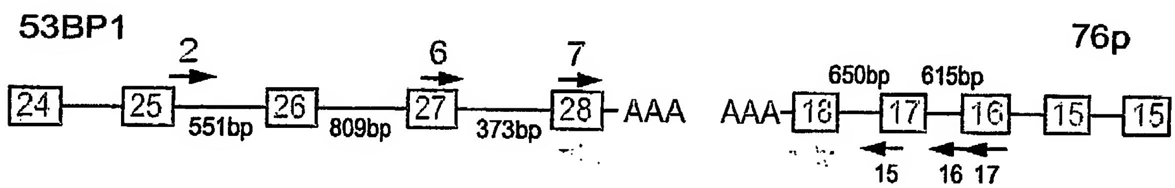


Fig. 10

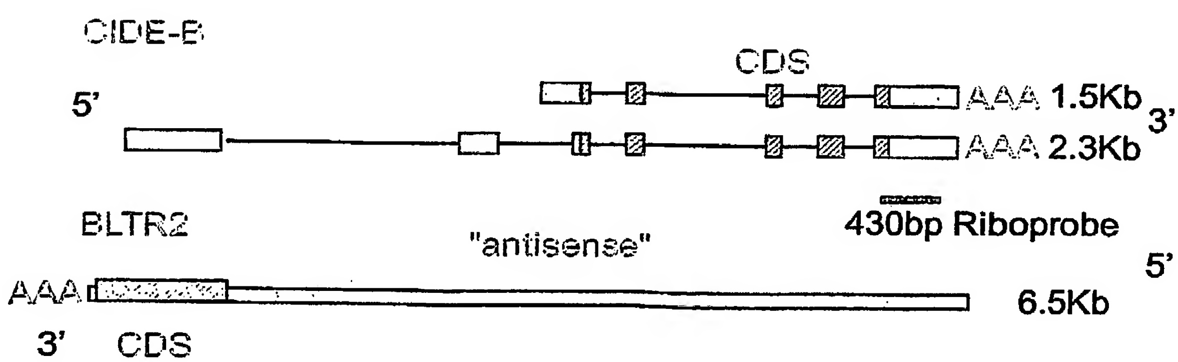
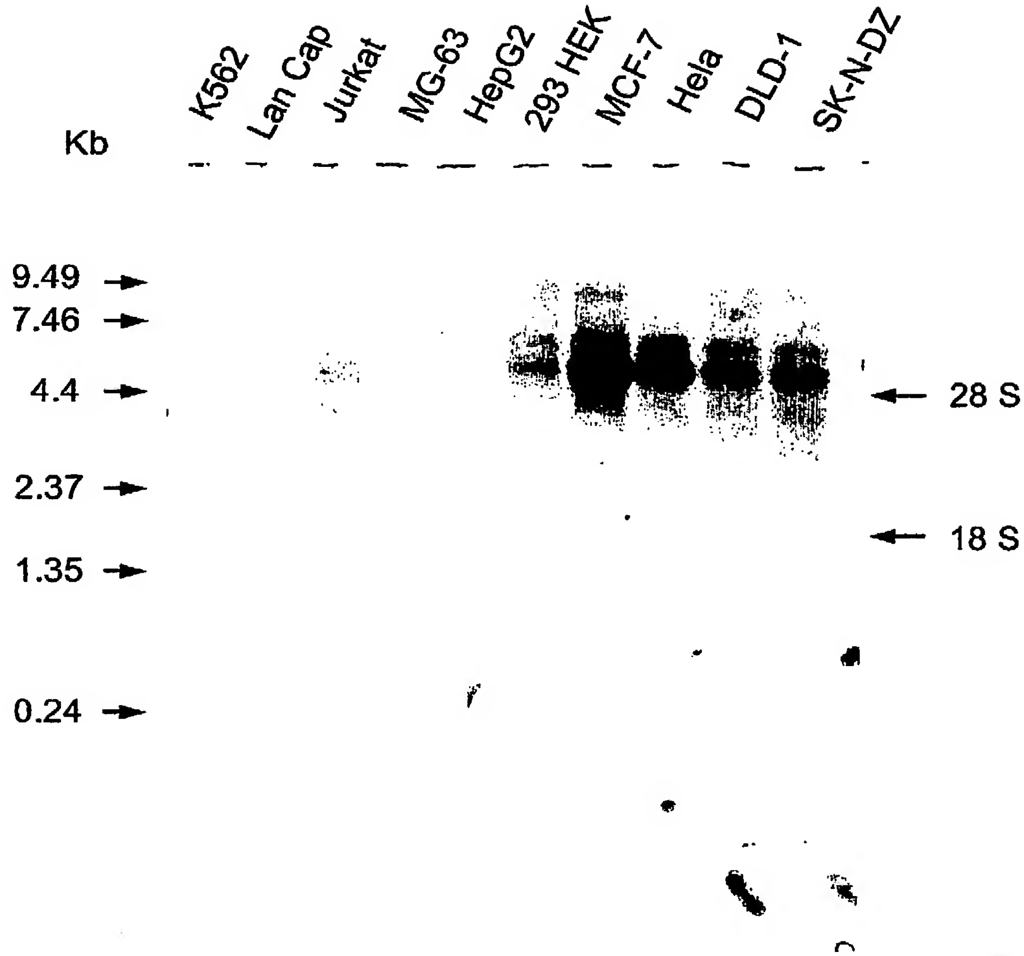


Fig. 11



K562
Lan Cap
Jurkat
MG-63
HepG2
293 HEK
MCF-7
Hela
DLD-1
SK-N-DZ
Kb
9.49
7.46
4.4
2.37
1.35
0.24
28 S
18 S

Fig. 12

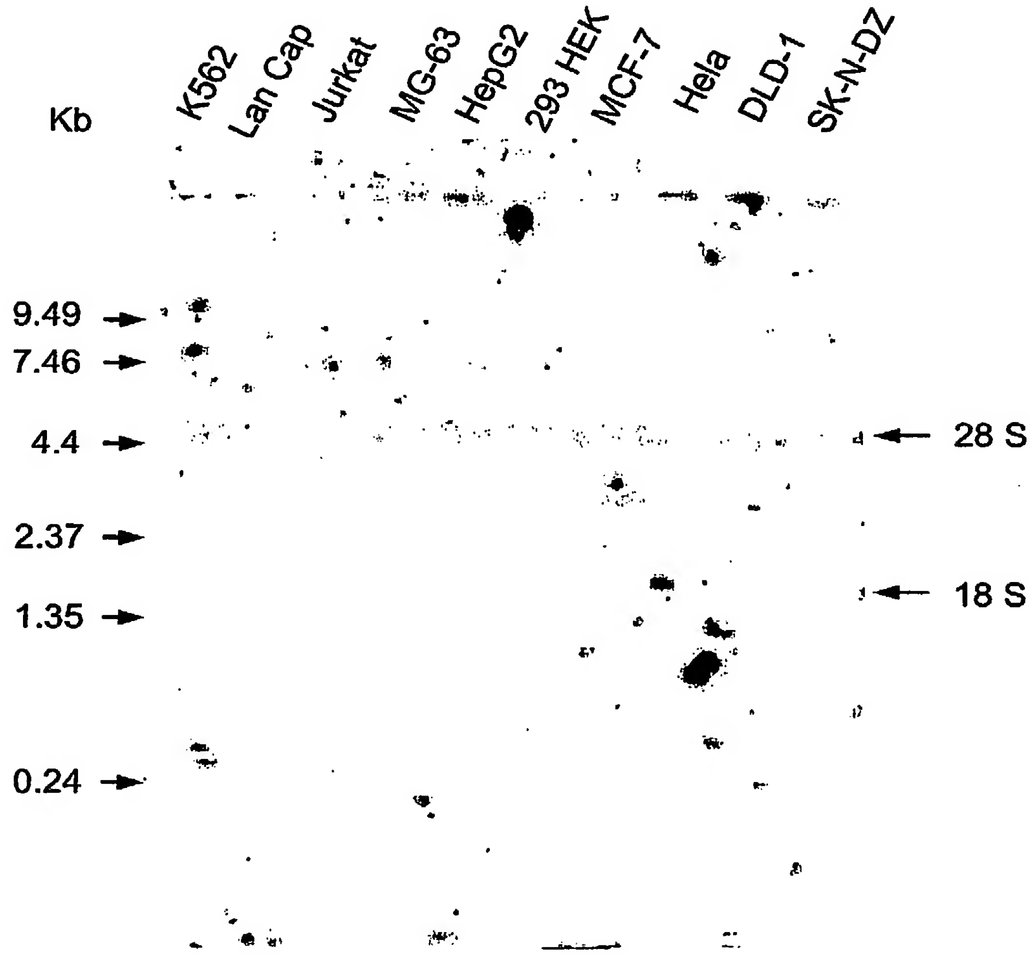

Kb
K562
Lan Cap
Jurkat
MG-63
HepG2
293 HEK
MCF-7
Hela
DLD-1
SK-N-DZ
9.49
7.46
4.4
2.37
1.35
0.24
28 S
18 S

Fig. 13

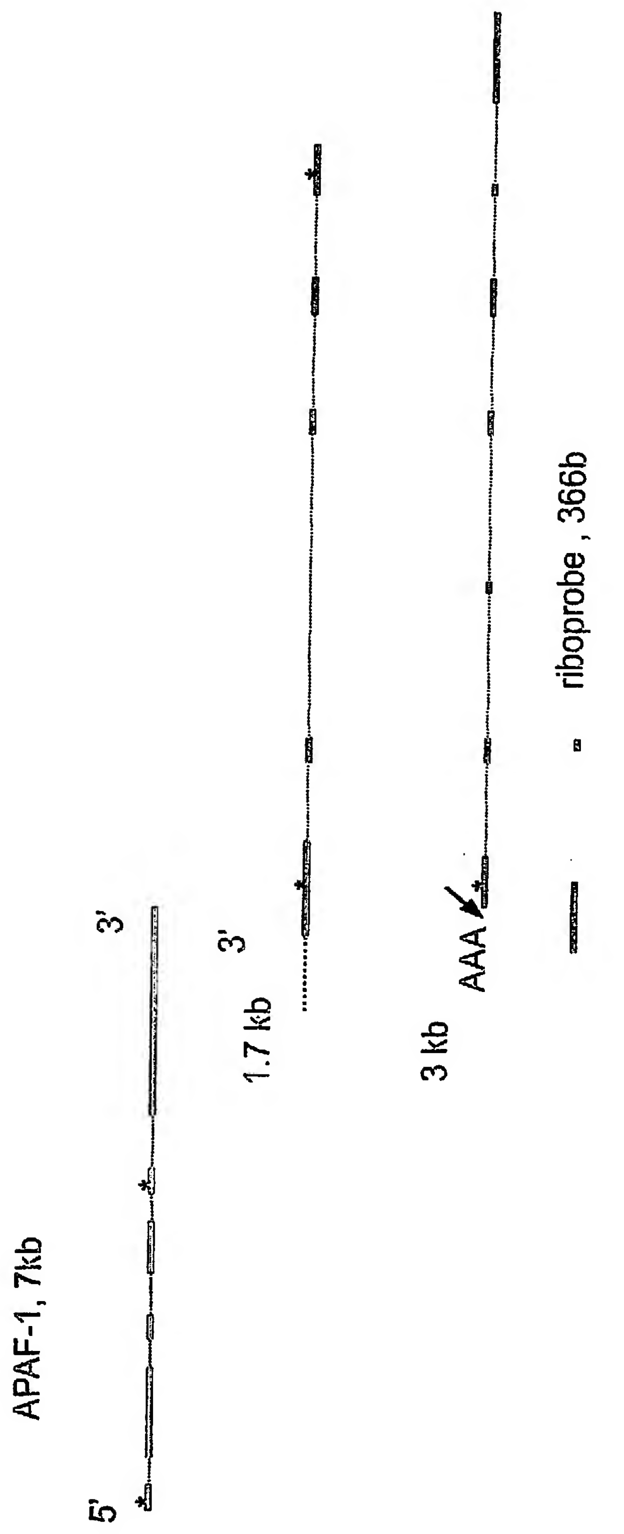


Fig. 14

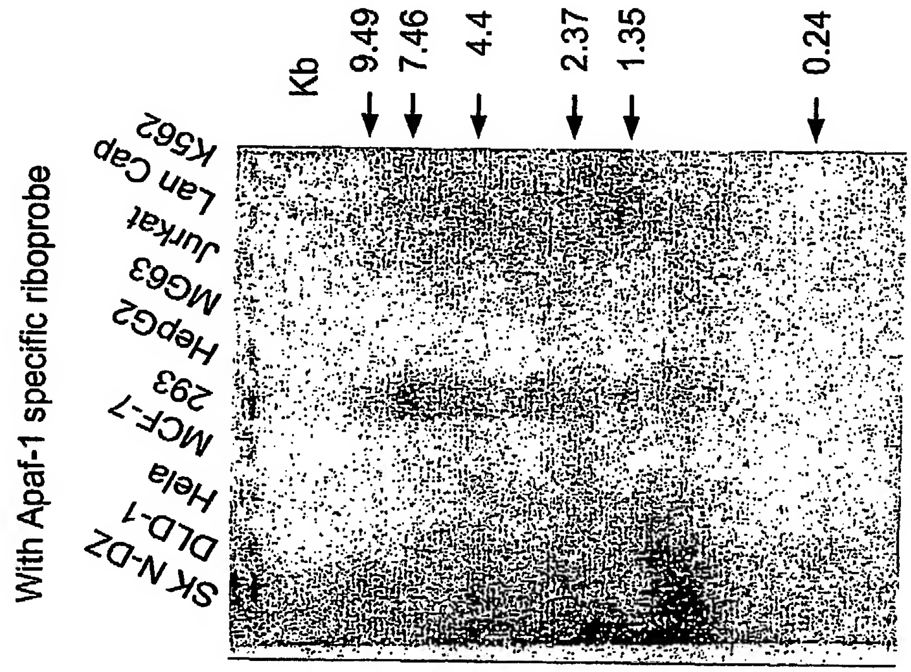


Fig. 15b

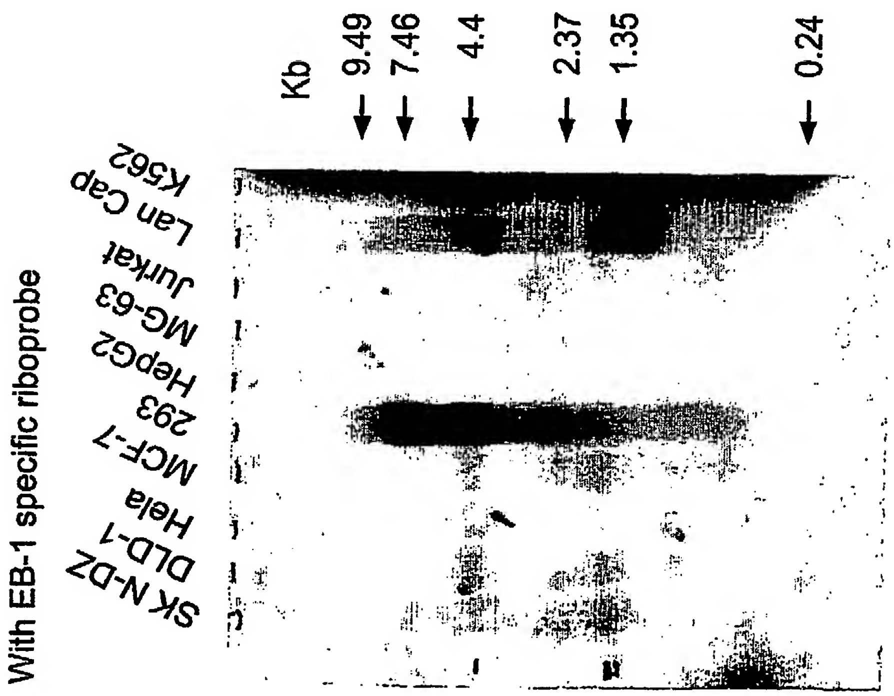


Fig. 15a

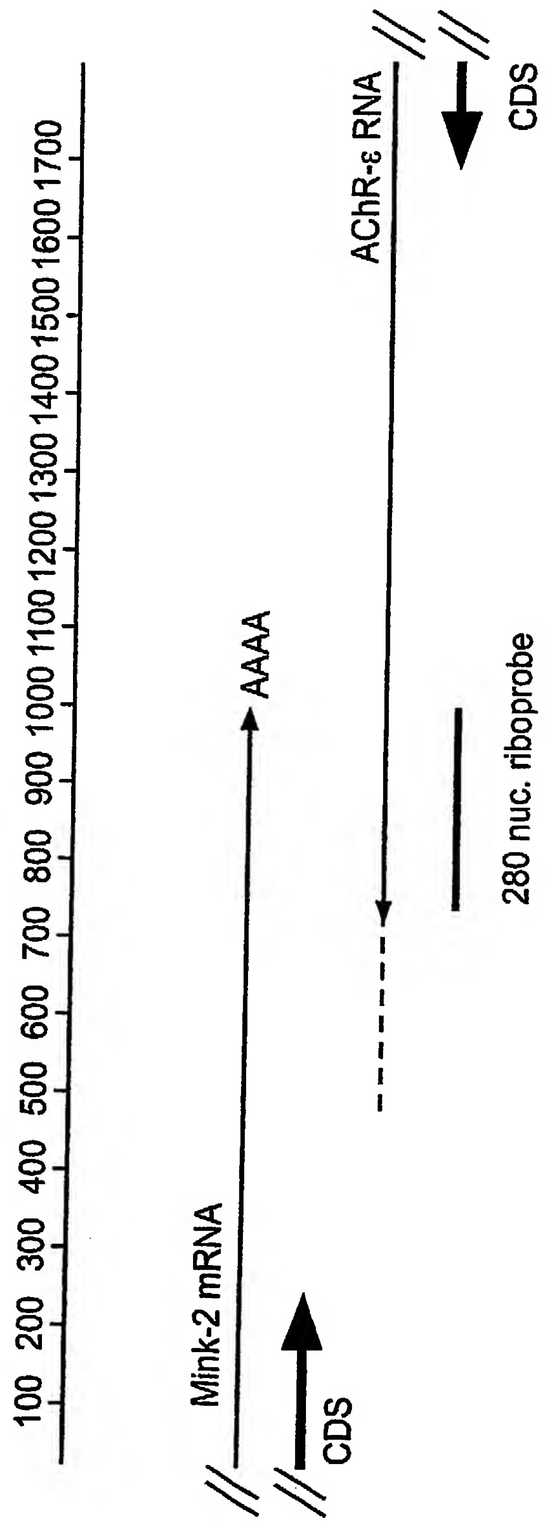

100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700
Mink-2 mRNA
AAAA
CDS
AChR-ε RNA
CDS
280 nuc. riboprobe

Fig. 16

Fig. 17a

AChR-ε

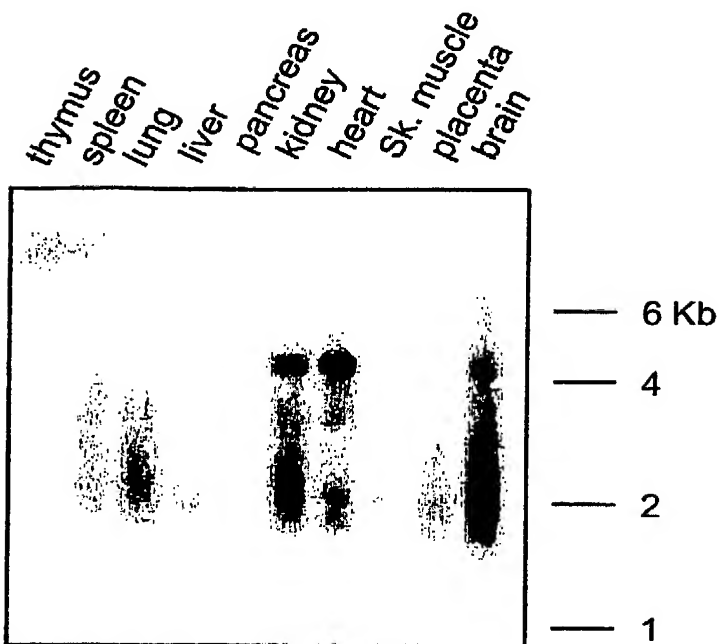
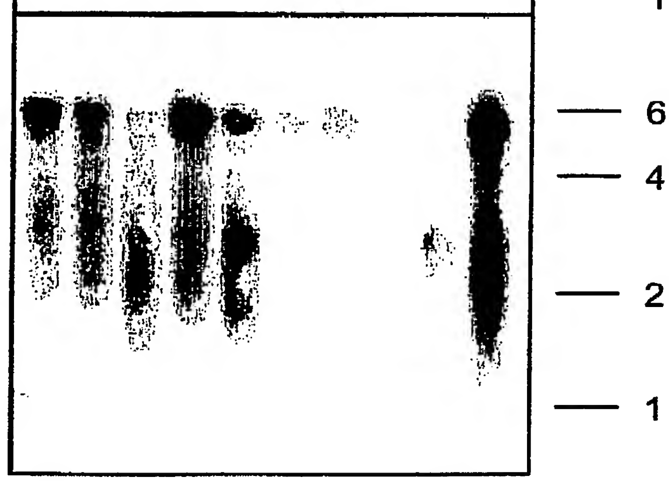


Fig. 17b

Mink



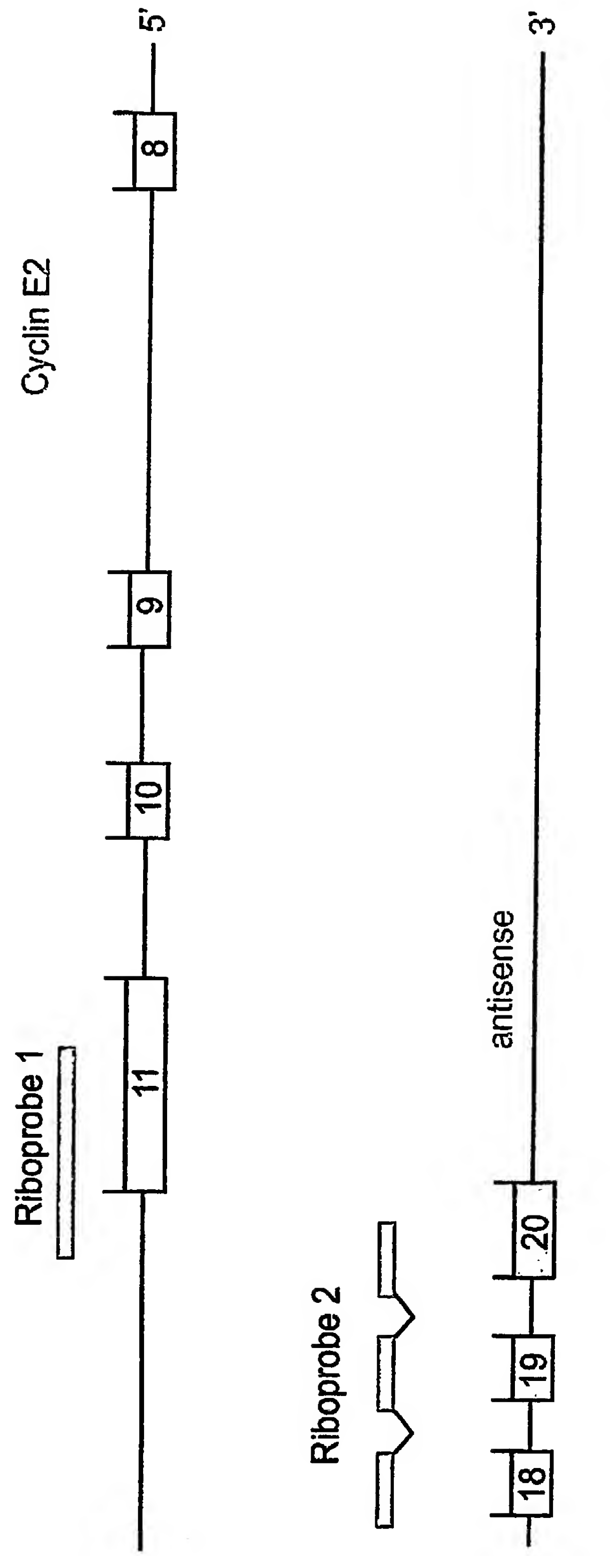

5'
3'
8
Cyclin E2
9
10
11
antisense
Riboprobe 1
20
Riboprobe 2
19
18

Fig. 18

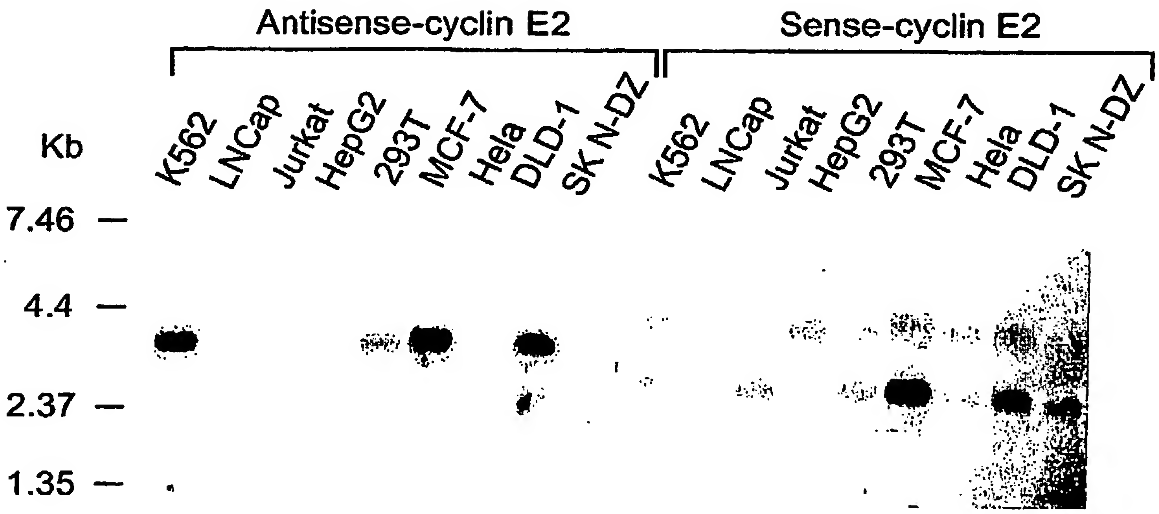


Fig. 19a Fig. 19b

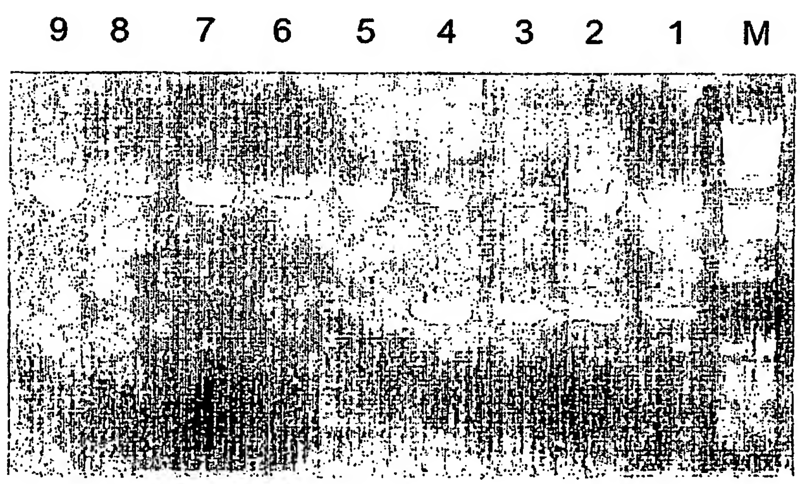


Fig. 20

Fig. 21a

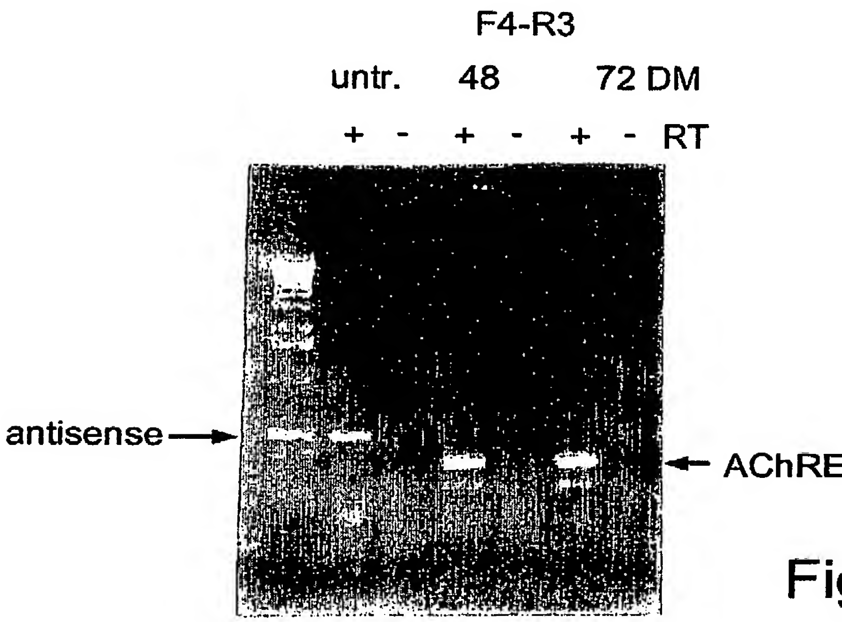


Fig. 21b

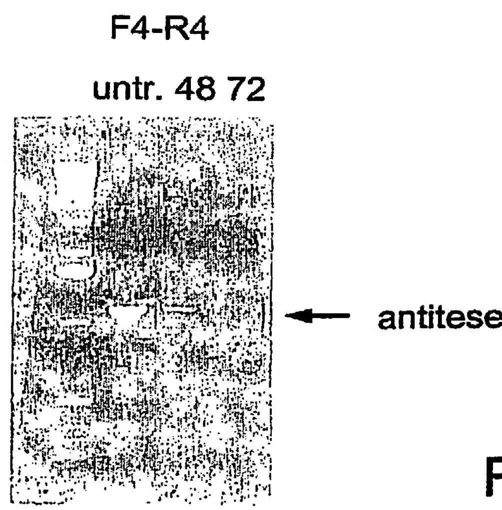


Fig. 21c

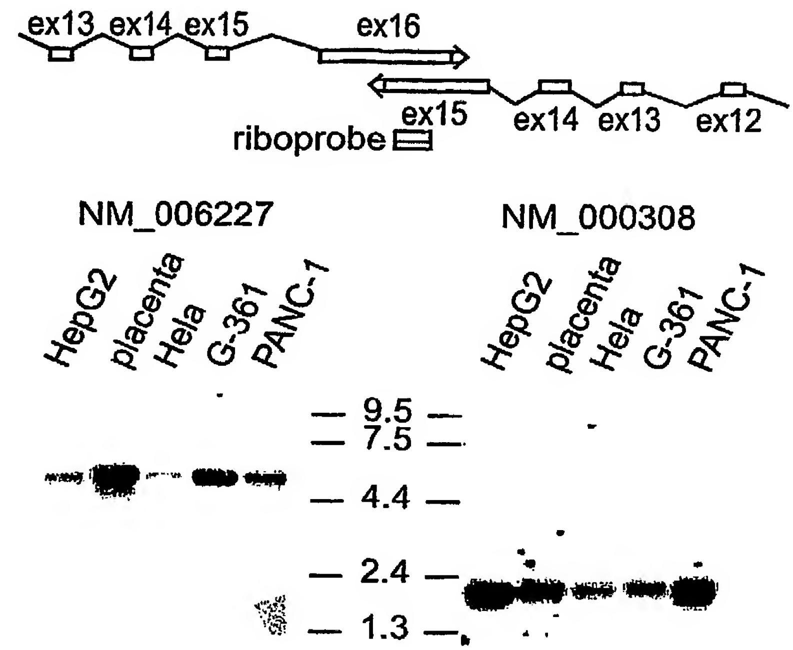


Fig. 22a

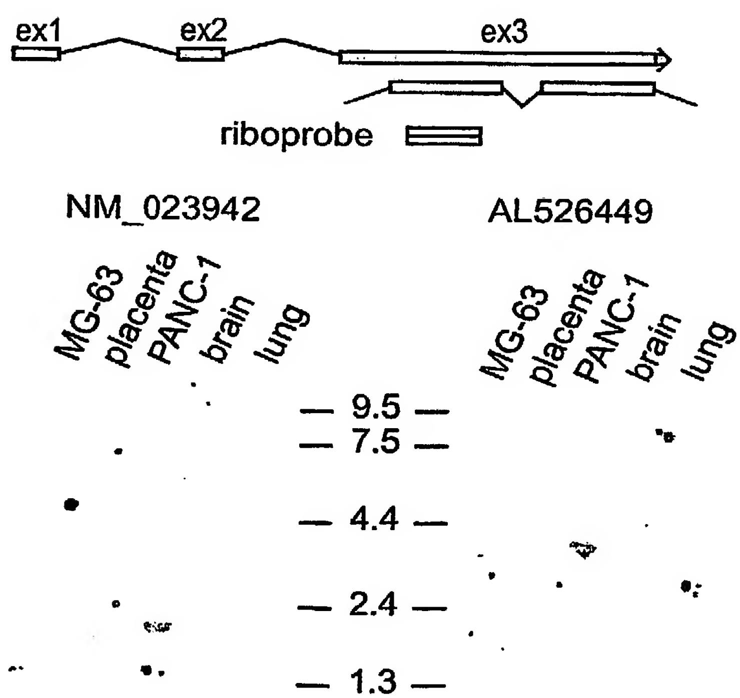


Fig. 22b

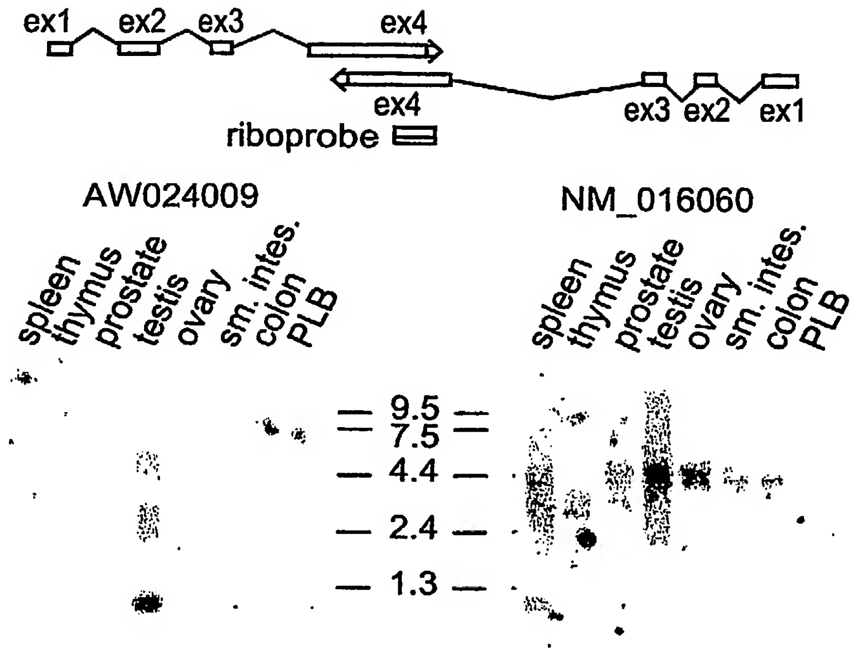


Fig. 22c

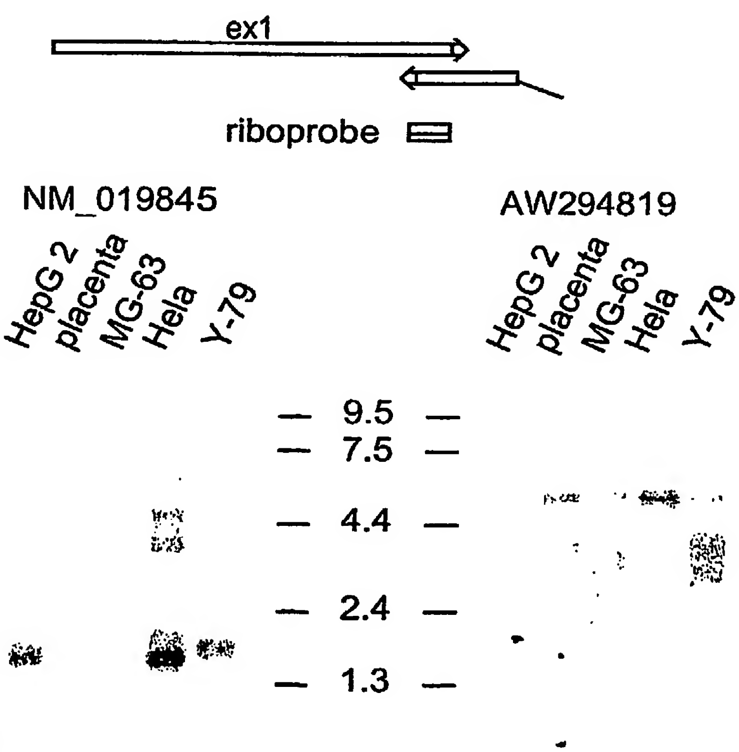


Fig. 22d

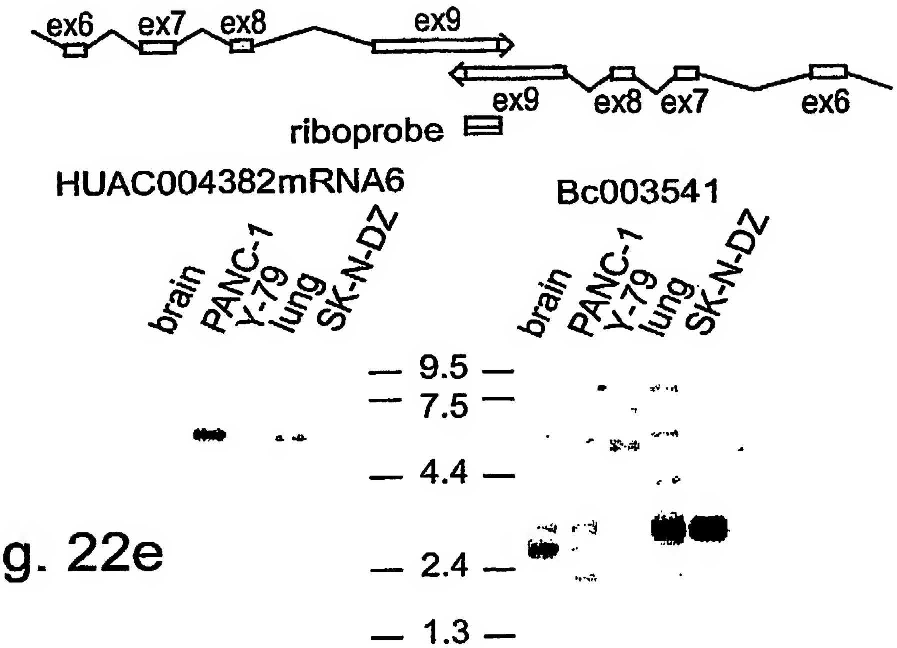


Fig. 22e

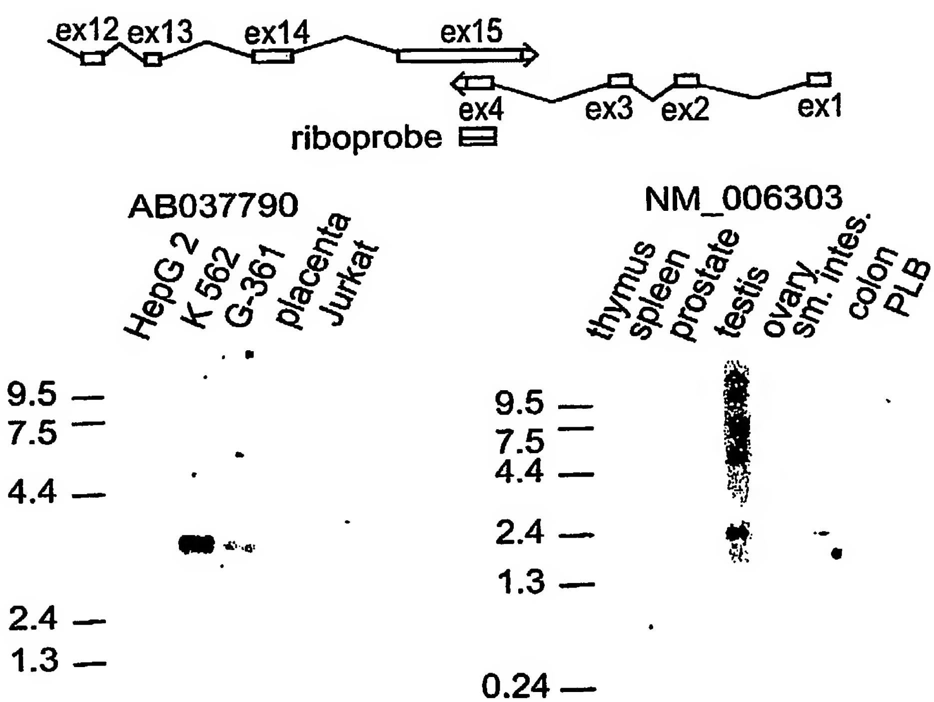


Fig. 22f

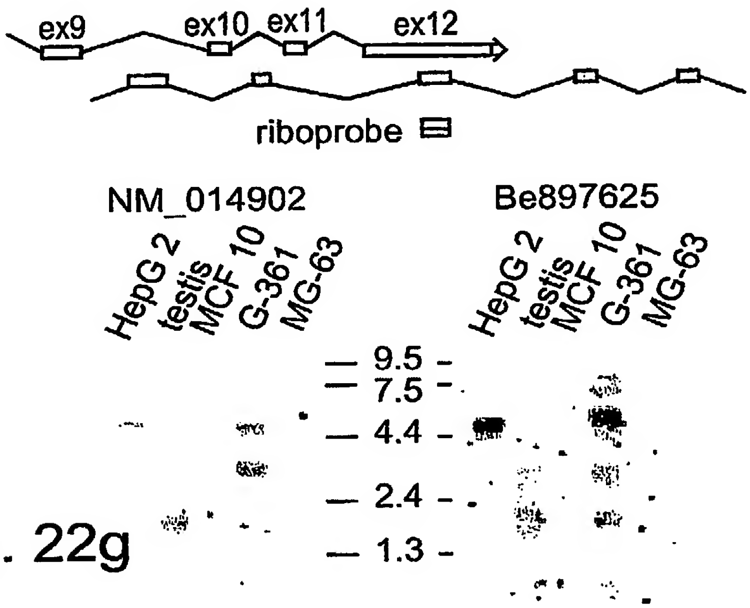


Fig. 22g

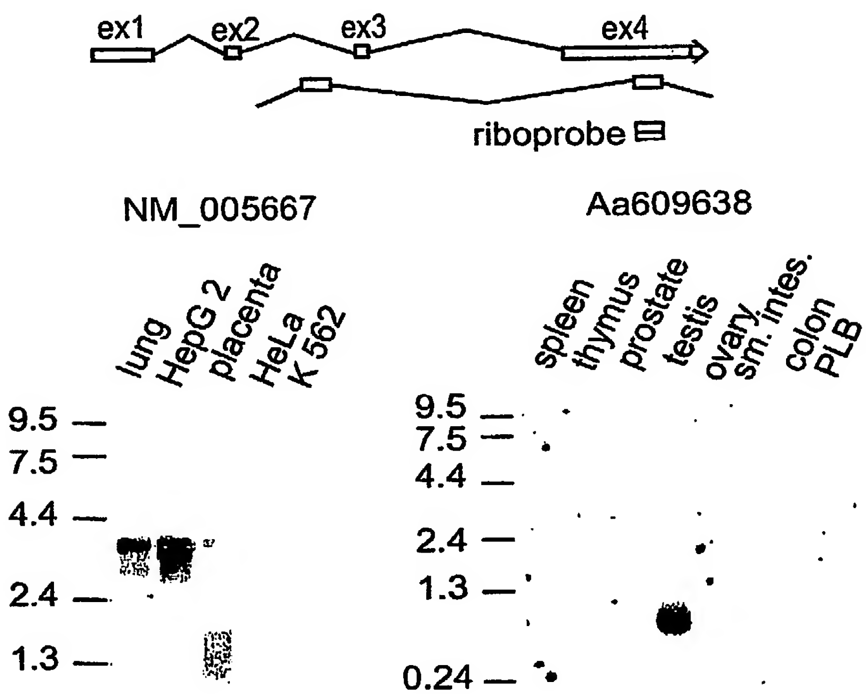


Fig. 22h

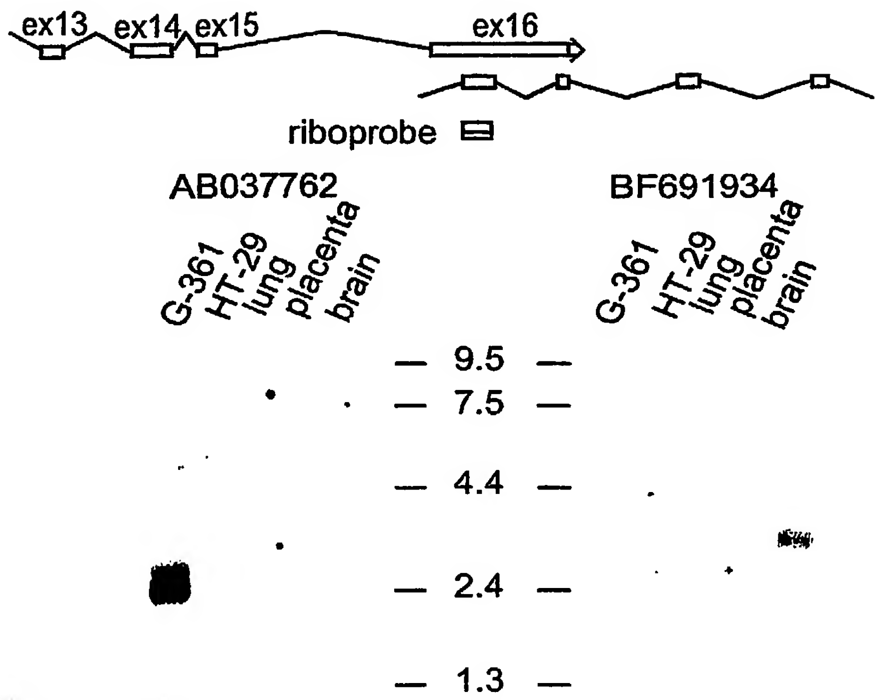


Fig. 22i

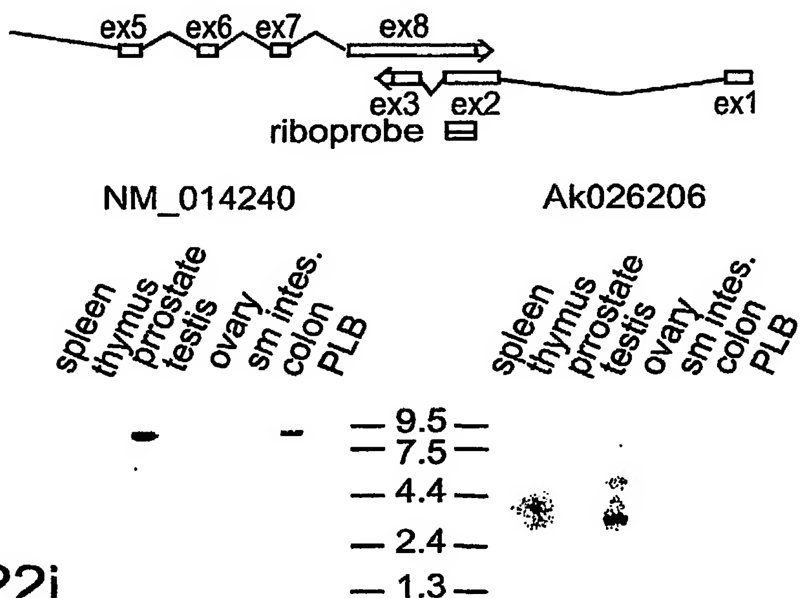


Fig. 22j

Probe	Probe pass thres hold	Anti sen se pair ver.	Actu al pool	Virtu al pool	Y79		ES-2		HEK -293		K-562		HT-29		SK-N DZ		MG-63		SNUI		NI564		HeLa		MCF10
					28	52b	29	53	30b	78	31b	55b	33b	56	34b	89	35	58	36b	79	45	63	46b	64	50b
Actin luM	+	+ CTRL	472	442	0.56	0.71	0.89	0.89	0.62	0.95	0.71	0.73	0.41	0.47	0.66	0.62	0.78	0.8	1.01	1.12	2.24	2.6	0.91	1.09	0.92
GAPDH 30um	+	+ CTRL	1847	1710	0.97	0.96	0.82	0.78	0.53	0.51	0.97	0.86	0.47	0.52	0.87	0.73	1.03	0.8	1.88	1.65	1.28	0.87	1.38	1.33	0.84
HUMM HBA123	+	+ CTRL	2245	2282	1.43	1.42	0.85	0.78	1.16	1.06	1.26	1.02	0.82	0.88	1.1	0.95	1.95	1.28	0.88	0.75	0.91	0.54	1.37	1.25	1.14
Known 10AC	+	VER.	378	426	1.82	1.49	0.65	0.54	0.94	0.91	1.32	0.82	1.03	0.66	1.61	1.57	2.34	2.37	0.8	0.65	1.06	0.55	1.15	0.9	0.79
Known 10C	+	VER.	95	103	1.04	1.27	0.78	0.84	-1	1.51	1	-1	1	1.16	-1	1.78	1	1	0.74	1.01	0.79	1	1.49	1	0.58
Known 9AC	+	VER.	852	959	2	1.84	0.81	0.69	1.07	1.26	1.56	1.14	1.11	0.85	1.55	1.27	1.14	0.83	0.76	0.63	0.89	0.49	1.28	1.25	0.77
260AC	+		154	197	1.34	1.95	0.35	1	0.9	1.79	0.84	1	0.71	1.42	0.87	1.35	0.47	1	0.59	1.3	0.78	1	1.14	1	1.31
260C	-		33	48	1	1	0.93	1	-1	1.8	-1	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1
261AC	+	VER.	344	357	-1	1	0.65	0.53	1.34	1.22	1.76	1.09	1	0.69	1.65	1	1.04	0.71	0.86	0.64	1.32	0.39	2.09	1.01	1
72C	-	IND. VER.	63	69	1.28	1.15	0.46	0.78	1	1	-1	1	1	2.32	1.14	1.78	-1	-1	1	1	1	1	1	1	0.6

Fig. 23

	Lung		Testis		Placenta		DU 145		T24		G-361		Jurkat		HepG2		MCF7		PANC-1	H1299	Brain
68b	70	74	71	75	72b	76	82	83	69	51	80	60	88	66	86b	67b	85	84	37	43	77
0.98	1.34	1.37	0.45	0.55	1.51	1	1.19	1.1	1.52	0.94	1.41	1.36	1.33	1	0.34	0.55	1.32	1.13	0.86	0.9	1.16
0.82	0.37	0.31	0.36	0.28	0.55	0.44	0.93	0.69	0.96	0.57	1.4	1.19	1.36	0.91	0.85	0.81	0.78	0.5	0.62	2.16	0.86
1.04	1.44	0.9	1.16	0.71	1.19	0.9	1.12	0.82	0.97	0.55	0.95	0.71	1.13	0.62	1.37	1.32	1.31	0.83	1.13	1.65	0.66
0.78	1.74	1.22	1.87	1.34	1.42	1.13	1.17	0.95	0.77	0.35	1.42	1	0.94	0.84	1.12	1.16	0.96	0.59	1.76	1.48	1.43
1.61	1.59	1	0.89	1	1	1	1	1	1	0.57	1	1	1	1	0.85	1.3	1	1	1.65	1.55	1
0.84	2.11	1.27	1.54	1.16	1.04	0.67	1.05	0.9	0.86	0.52	1.38	0.79	1.34	1.1	1.03	1.09	1.48	1.11	1.52	1.29	1.2
3.64	1.01	2.58	1.43	3.45	1.52	3.07	1.02	1	0.56	0.72	0.84	0.5	1	1	0.84	1.55	1.1	1.09	1	1.04	4.89
1	1	1	1	1	1	2.46	1	-1	1	-1	1	1	1	1	1	1.68	1	1	-1	-1	1
1	1	1	2.17	1.31	1.39	0.83	1.36	0.72	0.96	0.52	1	0.58	1.95	1	1.5	0.94	1.86	1	-1	-1	1
1.05	1	1	1	-1	1	1	1.02	1	1	0.99	1	1	-1	1.83	1	1	1	1	1	1	1

Fig. 23 (Cont.)

SEQUENCE LISTING

<110> Levanon Erez, et al.

<120> METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE TRANSCRIPTS AND METHODS AND KITS UTILIZING SAME

<130> 02/25320

<150> US 09/718,407

<151> 2000-11-24

<150> US 09/732,938

<151> 2000-12-11

<150> US 09/785,439

<151> 2001-02-20

<150> US 09/907,923

<151> 2001-07-18

<150> US 009/993,398

<151> 2001-11-06

<150> US 10/201,605

<151> 2002-7-24

<160> 44

<170> PatentIn version 3.1

<210> 1

<211> 190

<212> DNA

<213> Homo sapiens

<400> 1
ggacccagga tatgagcgga aaacactttc tctacttaga tacaactttt tcctgtgcgc 60

atgcctgtaa tcccagctac tcaggaggct gaggcaggag aatcccttga acccaggagg 120

cagaggttgc ggtgagccaa gatctcacca ttgcactcca gcctgggcaa taagaacaaa 180

actccgtctc 190

<210> 2

<211> 783

<212> DNA

<213> Homo sapiens

<400> 2
gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa 60

gaattcagtc cttgtggatg aactgtaaac agcacccttc ctctaagatg ccgaagatca 120

tagtttgtgg tttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc 180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc 240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa 300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa 360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta 420

atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca 480

gtgaaagatg aacactgggc cattttaaga tggaaacaag gtgggggttg gatagagagt 540

tatatgcagc ctcttttgca cctcgttggt atttgtaaga ccacattttt ttctccctag 600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag 660

tttcagattg cagtacagct atgtaactga gtaagacagg gagaaatatt aatccgtgag 720

tggctcccag taagaccatg gccaaataca tcctgaagta gaatatctgg aaaatttgag 780

att 783

<210> 3

<211> 1649

<212> DNA

<213> Homo sapiens

<400> 3
gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa 60

gaattcagtc cttgtggatg aactgtaaac agcacccttc ctctaagatg ccgaagatca 120

tagtttgtgg tttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc 180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc 240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa 300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa 360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta 420

atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca 480

gtgaaagatg aacactgggc cattttaaga tggaaacaag gtgggggttg gatagagagt 540

tatatgcagc ctcttttgca cctcgttggt atttgtaaga ccacattttt ttctccctag 600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag 660

tttcagattg cagtacagcc ttcaaatcat ctgggcccaa gttaaaacag aaggaattta 720

aaaaaaaaac acagtcactg tcttagaaga tgactcatat gctaagacag gtctgcctcc 780

ctgactcaga atgctgagtg actcctgaca ttattagttg gaatgggaag tgtaaggtca 840

agttggggtc tttacctgca tgacgaaacc acttcttgta atgacagact tttactgtgt 900

tggttagaat agccagtcct tggggagcct ctagtctgtt gtagctgaat gatttggaag 960

tgttctttca ctttttactt ttgtcctcag cattacctac atgaacttgt tatcaaaacc 1020

cttgtccagc acaacctctt ttatatgctg catcagttcc tgcagtacca cgtcctcagc 1080

gactccaaac ctttggcttg tctgctgtta tccctagaga gtttctatcc tcctgctcat 1140

cagctatctc tggacatgct gaaggtaact ctgatgtgtg aggttttaga ctatggaaac 1200

taactctgtt cctgttgttt gcactgacct ggacttctct cccttactgc tagcgacttt 1260

caacagcaaa tgatgaaata gtagaagttc tcctttccaa acaccaagtg ttagctgcct 1320

taaggtttat ccggggcatt ggtggccatg acaacatttc tgcacgaaaa tttttagatg 1380

ctgcaaagca gactgaagac aacatgcttt tctatacaat attccgcttt tttgaacagc 1440

gaaaccagcg tttgcgaggg agccccaatt tcacaccagg ggaacactgt gaagaacatg 1500

ttgctttttt caaacagatt tttggagacc aagctctaat gaggcctaca acattctgaa 1560

atcacttgct gttttttat ataaaaatgt gtacaaagtt aatttattgc attaataaag 1620

ctctttaaac tataaaatgt taaaaagtg 1649

<210> 4

<211> 1861

<212> DNA

<213> Homo sapiens

<400> 4

gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa 60

gaattcagtc cttgtggatg aactgtaaac agcacccttc ctctaagatg ccgaagatca 120

tagtttgtgg ttttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc 180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc 240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa 300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa 360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta 420

atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca 480

gtgaaagatg aacactgggc cattttaaga tggaaacaag gtgggggttg gatagagagt 540

tatatgcagc ctcttttgca cctcgttggt atttgtaaga ccacattttt ttctccctag 600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag 660

tttcagattg cagtacagcc ttcaaatcat ctgggcccaa gttaaaacag aaggaattta 720

aaaaaaaaac acagtcactg tcttagaaga tgactcatat gctaagacag gtctgcctcc 780

ctgactcaga atgctgagtg actcctgaca ttattagttg gaatgggaag tgtaaggtca 840

agttggggtc tttacctgca tgacgaaacc acttcttgta atgacagact tttactgtgt 900

tggttagaat agccagtcct tggggagcct ctagtctgtt gtagctgaat gatttggaag 960

tgttctttca ctttttactt ttgtcctcag cattacctac atgaacttgt tatcaaaacc 1020

cttgtccagc acaacctctt ttatatgctg catcagttcc tgcagtacca cgtcctcagc 1080

gactccaaac ctttggcttg tctgctgtta tccctagaga gtttctatcc tcctgctcat 1140

cagctatctc tggacatgct gaaggtaact ctgatgtgtg aggttttaga ctatggaaac 1200

taactctgtt cctgttgttt gcactgacct ggacttctct cccttactgc tagcgacttt 1260

caacagcaaa tgatgaaata gtagaagttc tcctttccaa acaccaagtg ttagctgcct 1320

taaggtttat ccggggcatt ggtggccatg acaacatttc tgcacgaaaa tttttagatg 1380

ctgcaaagca gactgaagac aacatgcttt tctatacaat attccgcttt tttgaacagc 1440

gaaaccagcg tttgcgaggg agccccaatt tcacaccagg tgagaatgca atgaaaagac 1500

ttggggtaac catagcctca aagagtagca gagggcactg gcagctggtg ggcgaggacc 1560

ctgggttagc atttttgtaa acaacacaat ttgataacag cccacctagc ccttggccca 1620

ttatttgtag tagagtgaat tcagtatact gacagaatct ggattatgct ctggaactca 1680

ccgaggaggt gtgttttgag tcaagacaca tttaggaccc agatcaggca cagcccatct 1740

cttatagcag atcttggaat atctcttaaa gccaggaata agacggcaaa tggtggctaa 1800

gggttttaaa gggtctgggg cttattaagg tttcagtttt atgaagtata cattggttga 1860

t 1861

<210> 5

<211> 214

<212> DNA

<213> Homo sapiens

<400> 5

gtaagggaac tttggcgact tagtgcgatc actgggagaa ttgtagagtc cactggagag 60

aaagaaaaat ggtcaaaaag agcccagaga gttcctgggg gaaaacacac cgcagcccag 120

acctattcat aactgcacag ctggtacttc cagaggcaca tgcaccaggg gcacgtggtt 180

ctctttgctg acaagattta ttaaaagaaa agag 214

<210> 6

<211> 1934

<212> DNA

<213> Homo sapiens

<400> 6

```
aagtcaacga aaggttccgt tgtccttgac cacgtattcc atcacgtaaa ccttgtggag      60
atagattatt ttgggctacg ttactgtgac agaagccatc agacgtattg gctggatcct     120
gcaaaaaccc ttgctgaaca caaagaactg atcaacactg gacctccata tactttgtat     180
tttggtatta aattctatgc tgaagatcca tgtaaactta aagaagaaat aaccagatat     240
cagtttttct tgcaggtgaa gcaagatgtc cttcagggcc gtctgccctg tcccgtcaac     300
actgctgctc agctgggagc gtatgccatc cagtcggagc ttggagatta tgacccatat     360
aaacatactg caggatatgt atctgagtac cggtttgttc ctgatcagaa ggaagaactt     420
gaagaagcca tagaaaggat tcataaaact ctaatgggtc agattccttc tgaggctgag     480
ctgaattact tgaggactgc caaatccctg gagatgtatg gcgttgacct ccatcccgtc     540
tatggagaaa acaagtctga gtatttctta ggattaactc cggttggtgt tgttgtgtac     600
aagaataaaa agcaagtggg gaagtatttc tggcctcgga ttacaaaggt tcacttcaag     660
gagactcaat ttgaactcag agtactggga aaagattgta acgaaacctc attctttttt     720
gaagctcgga gtaaaactgc ttgcaagcac ctctggaagt gcagtgtgga acatcataca     780
ttttttagaa tgccagaaaa tgaatccaat tcactgtcaa gaaaactcag caagtttgga     840
tccatacgtt ataagcaccg ctacagtggc aggacagctt tgcaaatgag ccgagatctt     900
tctattcagc ttccccggcc tgatcagaat gtgacaagaa gtcgaagcaa gacttaccct     960
aagcgaatag cacaaacaca gccagctgaa tcaaacacca tcagtaggat aactgcaaac    1020
atggaaaatg gagaaaatga aggaacaatt aaaattattg caccttcacc agtaaaaagc    1080
tttaagaaag caaagaatga aaatagccct gatacccaaa gaagcaaatc tcatgcaccg    1140
tgggaagaaa atggccccca gagtggactc tacaattctc ccagtgatcg cactaagtcg    1200
ccaaagttcc cttacacgcg tcgccgaaac ccctcctgtg gaagtgacaa tgattctgta    1260
cagcctgtga ggaggaggaa agcccataac agtggtgaag attcagatct taagcaaagg    1320
aggaggtcac gttcacgctg taacaccagc agtggtagtg aatcagaaaa ttctaataga    1380
gaacaccgga aaaagagaaa cagaatacgg caggagaatg atatggttga ttcagcgcct    1440
cagtgggaag ctgtattaag gagacaaaag gaaaaaaacc aagccgaccc caacagcagg    1500
cgatccagac acagatctcg ttcgagaagc cccgatatcc aagcaaaaga agagttatgg    1560
aagcacattc aaaaagaact tgtggatcca tccggattgt ccgaagaaca attaaaagag    1620
attccataca ctaaaataga gtgagtgcct ttcagaatct tctcaccaaa gctttattag    1680
tgcttgtgag taatccattc taattcttca attgtgttcc agacagtgct ttaatttgtc    1740
tttacatttt aaccaaaact aggtgacagt agcgaaagag gaagaaaagt gtgcattaaa    1800
gctacttatt ctacactata atcactatca tctcttatta gccacctctt tgtacttggt    1860
aggtacaagg gggcttttcc tgattaatgt cagttttaaa ataaattctt ttctgagatt    1920
```

ctcactgaaa aaat 1934

<210> 7

<211> 2353

<212> DNA

<213> Homo sapiens

<400> 7
aagtcaacga aaggttccgt tgtccttgac cacgtattcc atcacgtaaa ccttgtggag 60

atagattatt ttgggctacg ttactgtgac agaagccatc agacgtattg gctggatcct 120

gcaaaaaccc ttgctgaaca caaagaactg atcaacactg gacctccata tactttgtat 180

tttggtatta aattctatgc tgaagatcca tgtaaactta aagaagaaat aaccagatat 240

cagtttttct tgcaggtgaa gcaagatgtc cttcagggcc gtctgccctg tcccgtcaac 300

actgctgctc agctgggagc gtatgccatc cagtcggagc ttggagatta tgacccatat 360

aaacatactg caggatatgt atctgagtac cggtttgttc ctgatcagaa ggaagaactt 420

gaagaagcca tagaaaggat tcataaaact ctaatgggtc agattccttc tgaggctgag 480

ctgaattact tgaggactgc caaatccctg gagatgtatg gcgttgacct ccatcccgtc 540

tatggagaaa acaagtctga gtatttctta ggattaactc cggttggtgt tgttgtgtac 600

aagaataaaa agcaagtggg gaagtatttc tggcctcgga ttacaaaggt tcacttcaag 660

gagactcaat ttgaactcag agtactggga aaagattgta acgaaacctc attctttttt 720

gaagctcgga gtaaaactgc ttgcaagcac ctctggaagt gcagtgtgga acatcataca 780

ttttttagaa tgccagaaaa tgaatccaat tcactgtcaa gaaaactcag caagtttgga 840

tccatacgtt ataagcaccg ctacagtggc aggacagctt tgcaaatgag ccgagatctt 900

tctattcagc ttccccggcc tgatcagaat gtgacaagaa gtcgaagcaa gacttaccct 960

aagcgaatag cacaaacaca gccagctgaa tcaaacacca tcagtaggat aactgcaaac 1020

atggaaaatg gagaaaatga aggaacaatt aaaattattg caccttcacc agtaaaaagc 1080

tttaagaaag caaagaatga aaatagccct gatacccaaa gaagcaaatc tcatgcaccg 1140

tgggaagaaa atggccccca gagtggactc tacaattctc ccagtgatcg cactaagtcg 1200

ccaaagttcc cttacacgcg tcgccgaaac cctcctgtgg gaagtgacaa tgattctgta 1260

cagcctgtga ggaggaggaa agcccataac agtggtgaag attcagatct taagcaaagg 1320

aggaggtcac gttcacgctg taacaccagc agtggtagtg aatcagaaaa ttctaataga 1380

gaacaccgga aaaagagaaa cagaatacgg caggagaatg atatggttga ttcagcgcct 1440

cagtgggaag ctgtattaag gagacaaaag gaaaaaaacc aagccgaccc caacagcagg 1500

cgatccagac acagatctcg ttcgagaagc cccgatatcc aagcaaaaga agagttatgg 1560

aagcacattc aaaaagaact tgtggatcca tccggattgt ccgaagaaca attaaaagag 1620

attccataca ctaaaataga gacacaaggt gacccaatcc gcatcaggca ttctcattcg 1680

ccacgaagtt accgccagta tcgcaggtcc cagtgttcag atggggagcg atcagttctc 1740

tcggaagtga attcaaaaac agatcttgta ccaccacttc cggtgaccca ttcttcggat 1800

gctcagggtt ctggggatgc tacagttcat cagagaagaa atgggtctaa agatagcctg 1860

atggaagaaa aacctcagac atctacaaac aacctggctg gaaaacacac agcaaaaaca 1920

ataaaaacta tacaagcttc ccgcctcaag acagagactt gatcctgatg aagggtcaag 1980

ggtaggggtg ggaaggttgt gtgcgccact ggtacttttg aaactgtgaa ataggtatct 2040

taattcaaat ctcagacctg caagtatttc ttcagcatga gaaaatacat tatcttttgc 2100

ttcttttttt tttttttttg agatgttatc actctgtcgc ccaggctgga gtgcagcggc 2160

accgtgtcag ctcaccgcag cctccactta ctgggttaag cgattctcct gtctcaggct 2220

accgagcagc tgggattaca ggcgtgcacc acaacacccg gctaattctt tttgtatttt 2280

tagtagagac agggctttgc catgttggag gctggtctcg aactcctgac ctcaagtgat 2340

ccgcctgcct cag 2353

<210> 8

<211> 2500

<212> DNA

<213> Homo sapiens

<400> 8

gacatgggct gtttctgcgc tgttccggaa gaattttact gcgaagtttt gctcctggat 60

gaatccaagt taacccttac cacccagcag cagggcatca agaagtcaac gaaaggttcc 120

gttgtccttg accacgtatt ccatcacgta aaccttgtgg agatagatta ttttgggcta 180

cgttactgtg acagaagcca tcagacgtat tggctggatc ctgcaaaaac ccttgctgaa 240

cacaaagaac tgatcaacac tggacctcca tatactttgt attttggtat taaattctat 300

gctgaagatc catgtaaact taaagaagaa ataaccagat atcagttttt cttgcaggtg 360

aagcaagatg tccttcaggg ccgtctgccc tgtcccgtca acactgctgc tcagctggga 420

gcgtatgcca tccagtcgga gcttggagat tatgacccat ataaacatac tgcaggatat 480

gtatctgagt accggtttgt tcctgatcag aaggaagaac ttgaagaagc catagaaagg 540

attcataaaa ctctaatggg tcagattcct tctgaggctg agctgaatta cttgaggact 600

gccaaatccc tggagatgta tggcgttgac ctccatcccg tctatggaga aaacaagtct 660

gagtatttct taggattaac tccggttggt gttgttgtgt acaagaataa aaagcaagtg 720

gggaagtatt tctggcctcg gattacaaag gttcacttca aggagactca atttgaactc 780

agagtactgg gaaaagattg taacgaaacc tcattctttt ttgaagctcg gagtaaaact 840

gcttgcaagc acctctggaa gtgcagtgtg gaacatcata catttttag aatgccagaa 900

aatgaatcca attcactgtc aagaaaactc agcaagtttg gatccatacg ttataagcac 960

cgctacagtg gcaggacagc tttgcaaatg agccgagatc tttctattca gcttccccgg 1020

cctgatcaga atgtgacaag aagtcgaagc aagacttacc ctaagcgaat agcacaaaca 1080

cagccagctg aatcaaacac catcagtagg ataactgcaa acatggaaaa tggagaaaat 1140
gaaggaacaa ttaaaattat tgcaccttca ccagtaaaaa gctttaagaa agcaaagaat 1200
gaaaatagcc ctgataccca aagaagcaaa tctcatgcac cgtgggaaga aaatggcccc 1260
cagagtggac tctacaattc tcccagtgat cgcactaagt cgccaaagtt cccttacacg 1320
cgtcgccgaa acccctcctg tggaagtgac aatgattctg tacagcctgt gaggaggagg 1380
aaagcccata acagtggtga agattcagat cttaagcaaa ggaggaggtc acgttcacgc 1440
tgtaacacca gcagtggtag tgaatcagaa aattctaata gagaacaccg gaaaaagaga 1500
aacagaatac ggcaggagaa tgatatggtt gattcagcgc ctcagtggga agctgtatta 1560
aggagacaaa aggaaaaaaa ccaagccgac cccaacagca ggcgatccag acacagatct 1620
cgttcgagaa gccccgatat ccaagcaaaa gaagagttat ggaagcacat tcaaaaagaa 1680
cttgtggatc catccggatt gtccgaagaa caattaaaag agattccata cactaaaata 1740
gagtgagtgc ctttcagaat cttctcacca aagctttatt agtgcttgac acaaggtgac 1800
ccaatccgca tcaggcattc tcattcgcca cgaagttacc gccagtatcg caggtcccag 1860
tgttcagatg gggagcgatc agttctctcg gaagtgaatt caaaaacaga tcttgtacca 1920
ccacttccgg tgacccattc ttcggatgct cagggttctg ggatgctac agttcatcag 1980
agaagaaatg ggtctaaaga tagcctgatg gaagaaaaac ctcagacatc tacaaacaac 2040
ctggctggaa aacacacagc aaaaacaata aaaactatac aagcttcccg cctcaagaca 2100
gagacttgat cctgatgaag ggtcaagggt aggggtggga aggttgtgtg cgccactggt 2160
acttttgaaa ctgtgaaata ggtatcttaa ttcaaatctc agacctgcaa gtatttcttc 2220
agcatgagaa aatacattat cttttgcttc tttttttttt ttttttgaga tgttatcact 2280
ctgtcgccca ggctggagtg cagcggcacc gtgtcagctc accgcagcct ccacttactg 2340
ggttaagcga ttctcctgtc tcaggctacc gagcagctgg gattacaggc gtgcaccaca 2400
acacccggct aattcttttt gtatttttag tagagacagg gctttgccat gttggaggct 2460
ggtctcgaac tcctgacctc aagtgatccg cctgcctcag 2500

<210> 9

<211> 947

<212> DNA

<213> Homo sapiens

<400> 9
gaaagatgat actaggtcag gaaatagcat ttgaaagtca ttctcatctg gagggatgaa 60
gccaagataa ggcggaacca gggaaaagct ttaagaaagc aaagaatgaa aatagccctg 120
atacccaaag aagcaaatct catgcaccgt gggaagaaaa tggcccccag agtggactct 180
acaattctcc cagtgatcgc actaagtcgc caaagttccc ttacacgcgt cgccgaaacc 240
cctcctgtgg aagtgacaat gattctgtac agcctgtgag gaggaggaaa gcccattaac 300

agtggtgaag tttcagatct taaggcaaag ggagggaggt cacgttcacg ctgtaacacc 360

agcagtggta gtgaatcaga aaattctaat agagaacacc ggaaaaagag aaacagaata 420

cggcaggaga atgatatggt tgattcagcg cctcagtggg aagctgtatt aaggagacaa 480

aaggaaaaaa accaagccga ccccaacagc aggcgatcca gacacagatc tcgttcgaga 540

agccccgata tccaagcaaa agaagagtta tggaagcaca ttcaaaaaga acttgtggat 600

ccatccggat tgtccgaaga acaattaaaa gagattccat acactaaaat agagtgagtg 660

cctttcagaa tcttctcacc aaagctttat tagtgcttgt gagtaatcca ttctaattct 720

tcaattgtgt tccagacagt gctttaattt gtctttacat tttaaccaaa actaggtgac 780

agtagcgaaa gaggaagaaa agtgtgcatt aaagctactt attctacact ataatcacta 840

tcatctctta ttagccacct ctttgtactt ggtaggtaca agggggcttt tcctgattaa 900

tgtcagtttt aaaataaatt cttttctgag attctcactg aaaaaat 947

<210> 10

<211> 1366

<212> DNA

<213> Homo sapiens

<400> 10

gaaagatgat actaggtcag gaaatagcat ttgaaagtca ttctcatctg gagggatgaa 60

gccaagataa ggcggaacca gggaaaagct ttaagaaagc aaagaatgaa aatagccctg 120

atacccaaag aagcaaatct catgcaccgt gggaagaaaa tggcccccag agtggactct 180

acaattctcc cagtgatcgc actaagtcgc caaagttccc ttacacgcgt cgccgaaacc 240

cctcctgtgg aagtgacaat gattctgtac agcctgtgag gaggaggaaa gcccattaac 300

agtggtgaag tttcagatct taaggcaaag ggagggaggt cacgttcacg ctgtaacacc 360

agcagtggta gtgaatcaga aaattctaat agagaacacc ggaaaaagag aaacagaata 420

cggcaggaga atgatatggt tgattcagcg cctcagtggg aagctgtatt aaggagacaa 480

aaggaaaaaa accaagccga ccccaacagc aggcgatcca gacacagatc tcgttcgaga 540

agccccgata tccaagcaaa agaagagtta tggaagcaca ttcaaaaaga acttgtggat 600

ccatccggat tgtccgaaga acaattaaaa gagattccat acactaaaat agagacacaa 660

ggtgacccaa tccgcatcag gcattctcat tcgccacgaa gttaccgcca gtatcgcagg 720

tcccagtgtt cagatgggga gcgatcagtt ctctcggaag tgaattcaaa aacagatctt 780

gtaccaccac ttccggtgac ccattcttcg gatgctcagg gttctgggga tgctacagtt 840

catcagagaa gaaatgggtc taaagatagc ctgatggaag aaaaacctca gacatctaca 900

aacaacctgg ctggaaaaca cacagcaaaa acaataaaaa ctatacaagc ttcccgcctc 960

aagacagaga cttgatcctg atgaagggtc aagggtaggg gtgggaaggt tgtgtgcgcc 1020

actggtactt ttgaaactgt gaaataggta tcttaattca aatctcagac ctgcaagtat 1080

ttcttcagca tgagaaaata cattatcttt tgcttctttt tttttttttt ttgagatgtt 1140

atcactctgt cgcccaggct ggagtgcagc ggcaccgtgt cagctcaccg cagcctccac 1200
ttactgggtt aagcgattct cctgtctcag gctaccgagc agctgggatt acaggcgtgc 1260
accacaacac ccggctaatt ctttttgtat ttttagtaga gacagggctt tgccatgttg 1320
gaggctggtc tcgaactcct gacctcaagt gatccgcctg cctcag 1366

<210> 11
<211> 422
<212> DNA
<213> Homo sapiens

<400> 11
aatcttcata atccccatgt gtcaaaggag agaccaggtg gaggtaactg aatcatgggg 60
gtggtttccc caggctgttt ttgtgatagt gagtgagttc tcatgagatc tgatggtttt 120
ataaggggct cttccctcct ttgcttgtga agaaggtgcc ttttttcccc tttgccttct 180
gccatgattg taagtttcct gaggcctccc cagccaagct gaactgtgag tcaattaaac 240
ctcttttctt cgtaaattac ccagtcttga gcagttcttt acagcagtgt gaaaacagag 300
gaatacaccc atacatgcta ttctctgccc agaagccagg gggagcctgc cattaaaatg 360
aaagtcactc cttgactcag aaccctcaaa tagctttcat ctcacccaga aaaaaaagaa 420
aa 422

<210> 12
<211> 1532
<212> DNA
<213> Homo sapiens

<400> 12
aggtttctgc acaggaatat cgagagcgtc atgaacccga gctatagaga aaggagatga 60
ggcgtgagcc accgcacccg gctgacaagt gtccttctaa gaaacacaca gaggagaaga 120
cacagaagag gagagcacca tgtgatggta gacacagaaa ttggagttct acagccacaa 180
gccaaggaac tcctggagcc accaggagat ggaagatgca aagaactgat tttctctcag 240
agcctctgga gggagtgtgg ccctggtgac accttgattt tggacttctg gcctacagaa 300
ccatgcacac aggaggactt catttcccag gtctccttgc agtgaagttg aggccatgtg 360
actggtcttg ggccaatgga atgggtgcag aagggacaca gcccatttct agactcagcc 420
tgaaatgtcc tccataatcc ttactctttc tcccttcact cactggctgc aggaagctga 480
gaattatcct tggacttaca taaagcattt tggactttat gtaagtaaca acctgttgta 540
ttaagctact aagattttac ggttgtttgt taaatcagct aaccttaaac atcctaacaa 600
ctacaaatag aatacctgtt actgcataca taaaaataca aaaattagct ggatgtggtc 660

ccacctgtag tcccagctac tcgggaggct gaggcaggag aattgcttga acctgggagg 720

cggaggttgt ggtgagctga gatcgcacca ctgcactgca gcctgggcaa cagagcagga 780

ctctatctca aaaaaaaaac acaataaaca tttcttacct actgtagttt ttgtgggtca 840

ggaatctggg agcagcttag ttggatgatt tctgctcaca gtgttttatg aggttgcagt 900

caagatgttg gctggggctg tagtcatctg gagatttaac tacggctgga ggatccactt 960

caccatggtt cactcacctg gtgctggttg ctggcaggaa atttcagctc ttctcttata 1020

tggatctctt cacagattgc ttgagtgtcc tcaccgtatg gtgactggct tcctttacag 1080

aaatcagttg aagggaatgg gcaagtaaga aacagcaatg ctttttatga cctagtcctg 1140

aagttcccca ccattactta tgttcattgg aagccagttg ctaaggagag cctgcactca 1200

aagattgggg aaatagactt tatctttcaa agtgttgaag aatttgcaga cgtattttaa 1260

aaccaccaca caatccatca acacatcatg tcggctctat tcttgaaata gatccagaat 1320

ttgaccactt ttcaccatct ccattgctat tacccagatc taatcaacac catcacttgc 1380

ctggactaga gatttcctcc tcactgggct ctctgcttct atctttagcc cattgctatg 1440

atttggctgt gtccccaccc aaaatctcat cttgaattat aatcttcata atccccatgt 1500

gtcaaaggag agaccaggtg gaggtaactg aa 1532

<210> 13

<211> 1753

<212> DNA

<213> Homo sapiens

<400> 13

tttcttaggg tttttttttg agttggagcc tcgctctgtc ccccaggctg gagtgcagtg 60

atgtgatctc ggctcactgc aacctctgcc tcccaggttc aagtgattct cctgcctcag 120

cctccctagt agctgcgact acaggcatgt gccaccatgc ctggctaacg ttttgtattt 180

ttgagtagag acagggtttc accatgttgg ccaggctatt ctcgaactcc tgacctcaag 240

tgatccacct gcctcggctt cccaaagttt ctgggattac aggcgtgagc caccgcaccc 300

ggctgacaag tgtccttcta agaaacacac agaggagaag acacagaaga ggagagcacc 360

atgtgatggt agacacagaa attggagttc tacagccaca agccaaggaa ctcctggagc 420

caccaggaga tggaagatgc aaagaactga ttttctctca gagcctctgg agggagtgtg 480

gccctggtga caccttgatt ttggacttct ggcctacaga accatgcaca caggaggact 540

tcatttccca ggtctccttg cagtgaagtt gaggccatgt gactggtctt gggccaatgg 600

aatgggtgca gaagggacac agcccatttc tagactcagc ctgaaatgtc ctccataatc 660

cttactcttt ctcccttcac tcactggctg caggaagctg agaattatcc ttggacttac 720

ataaagcatt ttggacttta tgtaagtaac aacctgttgt attaagctac taagatttta 780

cggttgtttg ttaaatcagc taaccttaaa catcctaaca actacaaata gaatacctgt 840

tactgcatac ataaaaatac aaaaattagc tggatgtggt cccacctgta gtcccagcta 900

ctcgggaggc tgaggcagga gaattgcttg aacctgggag gcggaggttg tggtgagctg 960
agatcgcacc actgcactgc agcctgggca acagagcagg actctatctc aaaaaaaaaa 1020
cacaataaac atttcttacc tactgtagtt tttgtgggtc aggaatctgg gagcagctta 1080
gttggatgat ttctgctcac agtgttttat gaggttgcag tcaagatgtt ggctggggct 1140
gtagtcatct ggagatttaa ctacggctgg aggatccact tcaccatggt tcactcacct 1200
ggtgctggtt gctggcagga aatttcagct cttctcttat atggatctct tcacagattg 1260
cttgagtgtc ctcaccgtat ggtgactggc ttcctttaca gaaatcagtt gaagggaatg 1320
ggcaagtaag aaacagcaat gcttttatg acctagtcct gaagttcccc accattactt 1380
atgttcattg gaagccagtt gctaaggaga gcctgcactc aaagattggg gaaatagact 1440
ttatctttca aagtgttgaa gaatttgcag acgtatttta aaaccaccac acaatccatc 1500
aacacatcat gtcggctcta ttcttgaaat agatccagaa tttgaccact tttcaccatc 1560
tccattgcta ttacccagat ctaatcaaca ccatcacttg cctggactag agatttcctc 1620
ctcactgggc tctctgcttc tatctttagc ccattgctat gatttggctg tgtccccacc 1680
caaaatctca tcttgaatta taatcttcat aatccccatg tgtcaaagga gagaccaggt 1740
ggaggtaact gaa 1753

<210> 14

<211> 1832

<212> DNA

<213> Homo sapiens

<400> 14
gggttttgcg ggtataatta cattcaggat ctcaggatac tgcattatct gtgtgacccc 60
taaatctgat gacaagtgtc tgttttttgt ttttgttttt gagacagagc ctcgctctgt 120
cacccaggct ggagtgctgt ggtgtgatct cggctcactg caacctccgc ctcccaggtt 180
caagcaattc tctgcctcag cctcccgagt aaatgtgatt acaggcaggc gcctgccagc 240
acacccagct gattttagta tttttagtag agatggggtt tcaccatctt ggccaggctg 300
gtcttgaatt cctgacctcg tgatccaccc acttcagctt cccaaagttc tgggattaca 360
ggcgtgagcc accgcacccg gctgacaagt gtccttctaa gaaacacaca gaggagaaga 420
cacagaagag gagagcacca tgtgatggta gacacagaaa ttggagttct acagccacaa 480
gccaaggaac tcctggagcc accaggagat ggaagatgca aagaactgat tttctctcag 540
agcctctgga gggagtgtgg ccctggtgac accttgattt tggacttctg gcctacagaa 600
ccatgcacac aggaggactt catttcccag gtctccttgc agtgaagttg aggccatgtg 660
actggtcttg ggccaatgga atgggtgcag aagggacaca gcccatttct agactcagcc 720
tgaaatgtcc tccataatcc ttactctttc tcccttcact cactggctgc aggaagctga 780
gaattatcct tggacttaca taaagcattt tggactttat gtaagtaaca acctgttgta 840

ttaagctact aagattttac ggttgtttgt taaatcagct aaccttaaac atcctaacaa 900

ctacaaatag aatacctgtt actgcataca taaaaataca aaaattagct ggatgtggtc 960

ccacctgtag tcccagctac tcgggaggct gaggcaggag aattgcttga acctgggagg 1020

cggaggttgt ggtgagctga gatcgcacca ctgcactgca gcctgggcaa cagagcagga 1080

ctctatctca aaaaaaaaac acaataaaca tttcttacct actgtagttt ttgtgggtca 1140

ggaatctggg agcagcttag ttggatgatt tctgctcaca gtgttttatg aggttgcagt 1200

caagatgttg gctggggctg tagtcatctg gagatttaac tacggctgga ggatccactt 1260

caccatggtt cactcacctg gtgctggttg ctggcaggaa atttcagctc ttctcttata 1320

tggatctctt cacagattgc ttgagtgtcc tcaccgtatg gtgactggct tcctttacag 1380

aaatcagttg aagggaatgg gcaagtaaga aacagcaatg ctttttatga cctagtcctg 1440

aagttcccca ccattactta tgttcattgg aagccagttg ctaaggagag cctgcactca 1500

aagattgggg aaatagactt tatctttcaa agtgttgaag aatttgcaga cgtattttaa 1560

aaccaccaca caatccatca acacatcatg tcggctctat tcttgaaata gatccagaat 1620

ttgaccactt ttcaccatct ccattgctat tacccagatc taatcaacac catcacttgc 1680

ctggactaga gatttcctcc tcactgggct ctctgcttct atctttagcc cattgctatg 1740

atttggctgt gtccccaccc aaaatctcat cttgaattat aatcttcata atccccatgt 1800

gtcaaaggag agaccaggtg gaggtaactg aa 1832

<210> 15

<211> 10394

<212> DNA

<213> Homo sapiens

<400> 15

cgttgtttgg cgtgtttttt tttttgtttt ttgtcactgc ctgcctgggt cctgcccgag 60

gtctccatcc tcggtttccc tgtccttgcc ccgggccctg ggagtgctct ggaaggctgc 120

gcagtattgg aggggacaga atgaccttcc ggccttgagt ccctggggag cagatggacc 180

ctactggaag tcagttggat tcagatttct ctcagcaaga tactccttgc ctgataattg 240

aagattctca gcctgaaagc caggttctag aggatgattc tggttctcac ttcagtatgc 300

tatctcgaca ccttcctaat ctccagacgc acaaagaaaa tcctgtgttg gatgttgtgt 360

ccaatcctga acaaacagct ggagaagaac gaggagacgg taatagtggg ttcaatgaac 420

atttgaaaga aaacaaggtt gcagaccctg tggattcttc taacttggac acatgtggtt 480

ccatcagtca ggtcattgag cagttacctc agccaaacag gacaagcagt gttctgggaa 540

tgtcagtgga atctgctcct gctgtggagg aagagaaggg agaagagttg gaacagaagg 600

agaaagagaa ggaagaagat acttcaggca atactacaca ttcccttggt gctgaagata 660

ctgcctcatc acagttgggt tttggggttc tggaactctc ccagagccag gatgttgagg 720

aaaatactgt gccatatgaa gtggacaaag agcagctaca atcagtaacc accaactctg 780

```
gttataccag gctgtctgat gtggatgcta atactgcaat taagcatgaa gaacagtcca   840
acgaagatat ccccatagca gaacagtcca gcaaggacat ccctgtgaca gcacagccca   900
gtaaggatgt acatgttgta aaagagcaaa atccaccacc tgcaaggtca gaggacatgc   960
cttttagccc caaagcatct gttgctgcta tggaagcaaa agaacagttg tctgcacaag  1020
aacttatgga aagtggactg cagattcaga agtcaccaga gcctgaggtt ttgtcaactc  1080
aggaagactt gtttgaccag agcaataaaa cagtatcttc tgatggttgc tctactcctt  1140
caagggagga aggtgggtgt tctttggctt ccactcctgc caccactctg catctcctgc  1200
agctctctgg tcagaggtcc cttgttcagg acagtctttc cacgaattct tcagatcttg  1260
ttgctccttc tcctgatgct ttccgatcta ctccttttat cgttcctagc agtcccacag  1320
agcaagaagg gagacaagat aagccaatgg acacgtcagt gttatctgaa gaaggaggag  1380
agccttttca gaagaaactt caaagtggtg aaccagtgga gttagaaaac ccccctctcc  1440
tgcctgagtc cactgtatca ccacaagcct caacaccaat atctcagagc acaccagtct  1500
tccctcctgg gtcacttcct atcccatccc agcctcagtt ttctcatgac atttttattc  1560
cttccccaag tctggaagaa caatcaaatg atgggaagaa agatggagat atgcatagtt  1620
catctttgac agttgagtgt tctaaaactt cagagattga accaaagaat tcccctgagg  1680
atcttgggct atctttgaca ggggattctt gcaagttgat gctttctaca agtgaatata  1740
gtcagtcccc aaagatggag agcttgagtt ctcacagaat tgatgaagat ggagaaaaca  1800
cacagattga ggatacggaa cccatgtctc cagttctcaa ttctaaattt gttcctgctg  1860
aaaatgatag tatcctgatg aatccagcac aggatggtga agtacaactg agtcagaatg  1920
atgacaaaac aaagggagat gatacagaca ccagggatga cattagtatt ttagccactg  1980
gttgcaaggg cagagaagaa acggtagcag aagatgtttg tattgatctc acttgtgatt  2040
cggggagtca ggcagttccg tcaccagcta ctcgatctga ggcactttct agtgtgttag  2100
atcaggagga agctatggaa attaaagaac accatccaga ggaggggtct tcagggtctg  2160
aggtggaaga aatccctgag acaccttgtg aaagtcaagg agaggaactc aaagaagaaa  2220
atatggagag tgttccgttg cacctttctc tgactgaaac tcagtcccaa gggttgtgtc  2280
ttcaaaagga aatgccaaaa aaagaatgct cagaagctat ggaagttgaa accagtgtga  2340
ttagtattga ttcccctcaa aagttggcaa tacttgacca agaattggaa cataaggaac  2400
aggaagcttg ggaagaagct acttcagagg actccagtgt tgtcattgta gatgtgaaag  2460
agccatctcc cagagttgat gtttcttgtg aacctttgga gggagtggag aagtgctcag  2520
attcccagtc atgggaggat attgctccag aaatagaacc atgtgctgag aatagattag  2580
acaccaagga agaaaagagt gtagaatatg aaggagatct gaaatcaggg actgcagaaa  2640
cagaacctgt agagcaagat tcttcacagc cttccttacc tttagtgaga gcagatgatc  2700
ctttaagact tgaccaggag ttgcagcagc cccaaactca ggagaaaaca agtaattcat  2760
taacagaaga ctcaaaaatg gctaatgcaa agcagctaag ctcagatgca gaggcccaga  2820
agctggggaa gccctctgcc catgcctcac aaagcttctg tgaaagttct agtgaaaccc  2880
```

catttcattt cactttgcct aaagaaggtg atatcatccc accattgact ggtgcaaccc 2940
cacctcttat tgggcaccta aaattggagc ccaagagaca cagtactcct attggtatta 3000
gcaactatcc agaaagcacc atagcaacca gtgatgtcat gtctgaaagc atggtggaga 3060
cccatgatcc catacttggg agtggaaaag gggattctgg ggctgcccca gacgtggatg 3120
ataaattatg tctaagaatg aaactggtta gtcctgagac tgaggcgagt gaagagtctt 3180
tgcagttcaa cctggaaaag cctgcaactg gtgaaagaaa aaatggatct actgctgttg 3240
ctgagtctgt tgccagtccc cagaagacca tgtctgtgtt gagctgtatc tgtgaagcca 3300
ggcaagagaa tgaggctcga agtgaggatc cccccaccac acccatcagg gggaacttgc 3360
tccactttcc aagttctcaa ggagaagagg agaaagaaaa attggagggt gaccatacaa 3420
tcaggcagag tcaacagcct atgaagccca ttagtcctgt caaggacccct gtttctcctg 3480
cttcccagaa gatggtcata caagggccat ccagtcctca aggagaggca atggtgacag 3540
atgtgctaga agaccagaaa gaaggacgga gtactaataa ggaaaatcct agtaaggcct 3600
tgattgaaag gcccagccaa aataacatag gaatccaaac catggagtgt tccttgaggg 3660
tcccagaaac tgtttcagca gcaacccaga ctataaagaa tgtgtgtgag caggggacca 3720
gtacagtgga ccagaacttt ggaaagcaag atgccacagt tcagactgag agggggagtg 3780
gtgagaaacc agtcagtgct cctgggggatg atacagagtc gctccatagc cagggagaag 3840
aagagtttga tatgcctcag cctccacatg gccatgtctt acatcgtcac atgagaacaa 3900
tccgggaagt acgcacactt gtcactcgtg tcattacaga tgtgtattat gtggatggaa 3960
cagaagtaga aagaaaagta actgaggaga ctgaagagcc aattgtagag tgtcaggagt 4020
gtgaaactga agtttcccct tcacagactg gggctcctc aggtgacctg gggatatca 4080
gctccttctc ctccaaggca tccagcttac accgcacatc aagtgggaca agtctctcag 4140
ctatgcacag cagtggaagc tcagggaaag gagccggacc actcagaggg aaaaccagcg 4200
ggacagaacc cgcagatttt gccttaccca gctcccgagg aggcccagga aaactgagtc 4260
ctagaaaagg ggtcagtcag acagggacgc cagtgtgtga ggaggatggt gatgcaggcc 4320
ttggcatcag acagggaggg aaggctccag tcacgcctcg tgggcgtggg cgaaggggcc 4380
gcccaccttc tcggaccact ggaaccagag aaacagctgt gcctggcccc ttgggcatag 4440
aggacatttc acctaacttg tcaccagatg ataaatcctt cagccgtgtc gtgccccgag 4500
tgccagactc caccagacga acagatgtgg gtgctggtgc tttgcgtcgt agtgactctc 4560
cagaaattcc tttccaggct gctgctggcc cttctgatgg cttagatgcc tcctctccag 4620
gaaatagctt tgtagggctc cgtgttgtag ccaagtggtc atccaatggc tacttttact 4680
ctgggaaaat cacacgagat gtcggagctg ggaagtataa attgctcttt gatgatgggt 4740
acgaatgtga tgtgttgggc aaagacattc tgttatgtga ccccatcccg ctggacactg 4800
aagtgacggc cctctcggag gatgagtatt tcagtgcagg agtggtgaaa ggacatagga 4860
aggagtctgg ggaactgtac tacagcattg aaaaagaagg ccaaagaaag tggtataagc 4920
gaatggctgt catcctgtcc ttggagcaag gaaacagact gagagagcag tatgggcttg 4980
gcccctatga agcagtaaca cctcttacaa aggcagcaga tatcagctta gacaatttgg 5040

```
tggaagggaa gcggaaacgg cgcagtaacg tcagctcccc agccacccct actgcctcca   5100
gtagcagcag cacaacccct acccgaaaga tcacagaaag tcctcgtgcc tccatgggag   5160
ttctctcagg caaaagaaaa cttatcactt ctgaagagga acggtcccct gccaagcgag   5220
gtcgcaagtc tgccacagta aaacctggtg cagtaggggc aggagagttt gtgagcccct   5280
gtgagagtgg agacaacacc ggtgaaccct ctgccctgga agagcagaga gggcctttgc   5340
ctctcaacaa gaccttgttt ctgggctacg catttctcct taccatggcc acaaccagtg   5400
acaagttggc cagccgctcc aaactgccag atggtcctac aggaagcagt gaagaagagg   5460
aggaattttt ggaaattcct cctttcaaca agcagtatac agaatcccag cttcgagcag   5520
gagctggcta tatccttgaa gatttcaatg aagcccagtg taacacagct taccagtgtc   5580
ttctaattgc ggatcagcat tgtcgaaccc ggaagtactt cctgtgcctt gccagtggga   5640
ttccttgtgt gtctcatgtc tgggtccatg atagttgcca tgccaaccag ctccagaact   5700
accgtaatta tctgttgcca gctgggtaca gccttgagga gcaaagaatt ctggactggc   5760
aaccccgtga aaatcctttc cagaatctga aggtactctt ggtatcagac caacagcaga   5820
acttcctgga gctctggtct gagatcctca tgactggtgg tgcagcctct gtgaagcagc   5880
accattcaag tgcccataac aaagatattg ctttaggggt atttgatgtg gtggtgacgg   5940
acccctcatg cccagcctcg gtgctgaagt gtgctgaagc attgcagctg cctgtggtgt   6000
cacaagagtg ggtgatccag tgcctcattg ttggggagag aattggattc aagcagcatc   6060
caaaatataa acacgattat gtttctcact aaagatactt ggtcttactg gttttattcc   6120
ctgctatcgt ggagattgtg ttttaaccag gttttaaatg tgtcttgtgt gtaactggat   6180
tccttgcatg gatcttgtat atagttttat ttgctgaact tttatgataa aataaatgtt   6240
gaatctcttt ggttgtagta actgggattt cttcatctgt ttttttgagc ttaatctcag   6300
aacaaatgac aagacatagt actttctctg agtctttcaa caggcttatt cacttacgga   6360
ggacagctca ccaaggaaat tgaaaagtta agagtgaact ttattctgtg gcatcattcc   6420
caaaaggtta ttccagggtg tctaaaatgc tatgcttgca gaaactcagt ttaaggtagg   6480
tgaaggccca gattaacagt tgtgccaaaa gttgagtgga attgggcaca gctctgtttc   6540
ctgacagtta aaaaagacct catgctctct ctctgagctg agatcacagc tcacctgtgg   6600
gtactcccca actcttagag ctaaagggag aacgaaagga ccaactgcca tgaagggaca   6660
gtgaccataa gcttgatgga atgaccttcc gtaagataaa catgggaagc acaagtgaga   6720
acacctggaa atgttacacg ttctagtcaa agacccaata ttattattat tattattgtc   6780
acaatagctg gaagcagttc cttcccttcc tctggcatca ctgatccctg catggcttct   6840
cattctctaa agcaggggtc aacaaggttt ttttctgtaa agggtcaaag agtaaatatt   6900
tcaggctttg tgggccattt gatccatcac aactactcgc ctttgctgtg agggcatgaa   6960
agcaaccata gacaatgagt aaacaaatgg gcacggctgt gtttcagtaa aactgtacaa   7020
aaacagacag caggccatag tttgccagct cctgctccag agacagcagt ggaaagggtg   7080
atctttagtt gataatagca gggaataagt tgtcagagct tcccagtgtg tgtagaatat   7140
```

gtagtgatga aaaccagatg cagtgactat aacctgatgc cagaacactg cattcttttt 7200
cagtttggag ggcgttgttc agtgaatatt tctttttact tacactgata tgaatattga 7260
ttaccagtga tggctgggcc atattaagat aacttcaacc cctatggttt gtgtaagatg 7320
ggtaattggg cctgcaatct tcagtattta aaaatctaac aacttgatct caattttttc 7380
ttaaggacct ttttcttgga gaataatact tttttttttt tttttttttt tgagacggaa 7440
tttcgctctt gttgcccagg ctggaatgca atggcacaat ctcagctcac tgcagcgtct 7500
gcttcccagg ttcaagcaat tctcctgtct cagcctcctg agtagctggg attacaggca 7560
catgccacca cacctggcta atttttgtat ttttagtaga atcgaggttt catcatgttg 7620
gtcaggctgg tctcaaactc ctgacttcag gtgatccgcc cgcctcggcc tcccaaagtg 7680
ctgggattac aggtgtgagc caccatgccc ggcctaagaa atacttttaa gtatattttc 7740
attagctaga attgcccaat ctgtgtaggt ataaattact tggtataggg agagagaaag 7800
cctatcttac ctgttgcttt cttacttggt ggtaacatcc agcagttagt ctatttataa 7860
acataattac tttttcacat atgaaccata aaatatttaa ctttctgctc tatattgttt 7920
gtttaccgct gtatctccca cagcttgaac agtaccaagg tacgtagtag gtgctcaata 7980
aatgactatt gaataaatga acatatccaa caaatgttct caatgtaaag gatcagagat 8040
gccacatgtt ctccttgatg ggagagaccc ttccacatgg gaatgatggg aaggagttgt 8100
actcctggat gttcagtaac tgcttctagg agaaaaggta gagtcctatc actaagccgc 8160
agatatttat ttgtgtgtgg ctagaatggg atgttttgaa tcttctgtta caaccttggg 8220
aacgtggctg ttatttcaat ttatgagcca gaaattttca catcccgaaa ctacaaaaga 8280
gaaaaagagc cttattaagt gtcatgcttt cccaagacta ccttcaaaga aatatgaatc 8340
aggataacct gtgatctaaa taatgtcatc ttaaaactga agagtttctt ttgactcttc 8400
tgctacaata gcttagaaaa aaatctgctt gcagacattt tagagagaaa ggacaatgaa 8460
gtgattttct gaatgggaat gacagacctc tgggaagcca gctaccactg aatctcggta 8520
tcagtttttt ttaaagttta gagttagaag gggtggtcgc ctcctttcac agatgcggaa 8580
gctagggacc agcaaggcgg ggtgcccacg gctgcacagc tagttcatat cagaattggg 8640
agtggaaggc ccactgcctc ccagcatagc aatacataac ctagcaaagg acttaacacc 8700
tatctcactg tcaggttttt tagtatttta tgatgatgat gacttctact agaaaataac 8760
ctccattaaa attattaaag atggtcacac ctctatctct aagccttact tataaaatga 8820
gggtatttgg actaaagtct tcttccagtt ctagaattct acaactcatt aaaaagccac 8880
cttaaaaagt ctactgagtt acccaagggt tgctcctacc tgcccagagt tccaccagcc 8940
tgggtatagt atttgttata atctagtcgt aacagtagtt gagccaaatc tgagttgatc 9000
tgatgattcc gaacactgga gagaatcttg aacaggagtg aagactggcg gctaaagccc 9060
tgcagagaga aggactcagc tgtcattcca cttcagctca ccaactctcc atatggagga 9120
tgggggcgga gggaggagtt tcttggaaaa gccttgttca aaattctaca gaaccacctg 9180
gccttcccac attcctattt ctattagttc ctaaaatgac ttgtaccaaa tccatacatg 9240
catgacttcc tatgaaagta ctctttcatc agtaggaatt tagtagctgg tttccagtta 9300

atgtattttg tcaagtactg gggttgggga gaacccgttt tgattacaag cagataatta 9360

tctcagtgag atgggggtta gttcaaggaa gtaaggaggg gggaggatgt gaggaagtta 9420

gaacaaccca atgcttattt gatgggctga ataaactatt caggactgaa ctatttttga 9480

gcactgtgag gtggcacagt aattacctgc ttcaaaatca actgatacca acatttttat 9540

ctttgtatct tatctctgta cgtgtgtgta ttgaggaaat gctttactga ctcagaggaa 9600

agatcatgaa ttctccattg gcaaaaccac ctctgtcctt tcggcaaggc tgcatacttc 9660

caggcagacg caccttcacg agaatgctca gctgggcggc tccacgctca tccagtgggc 9720

ctaggttctg actgaccagc gaacaaaaac tgtgacagag atctaggatt tcattcaggc 9780

agtgaaacac ctacccggga aacagagttg gcattaggaa aggaaggaag gtacatccat 9840

gaagttaaag tgttaggaga acagtctgat taatagctga tctaattaat agctgacctc 9900

ccaaatctga caggatagac actgccacgt gcaaggcctg ccagcccctc agacgcacaa 9960

aatgcgtaaa acaaatgcat cctttcctgg ctaagcgagt attactctct tagccctgca 10020

ccaaacctcc aatctagcca catttaactc ttcatttctt agacccgcag agtgtcttcc 10080

tgcctctgag ctgtgagtgt tgttcccttt gcccgggatg ctcttgtttt taataccagt 10140

tcaagtccca ctctctcagt gaagcactcc cttccccact atagccttta gtgaaccctc 10200

gtttcttgct tctttattat ctgtactgtt gtccacttgg caattgttca ggcctctgtg 10260

ttgttactga tttttgtatg tatatatata tatatgtctt gtttttccaa ctagattgtg 10320

agctccttaa gggcagagcc atgaattata cctctttgta tccccagtgc cttgcataca 10380

gtaagcactc aata 10394

<210> 16

<211> 6837

<212> DNA

<213> Homo sapiens

<400> 16

agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg 60

gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgacccccttc ccagcttccc 120

gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag 180

gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg 240

aggagggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct 300

cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca 360

ggtatcgcag gacttcccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg 420

ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt 480

gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct 540

gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga 600

tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct 660

agaccagttc cagcttcttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca 720

aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg gggggttgcc 780

tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catggggtca tgtataaaca 840

gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa 900

acaggggcca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg 960

cattggggga ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga 1020

agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc 1080

ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt 1140

tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac 1200

ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga 1260

acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga 1320

agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg 1380

agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc 1440

agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga 1500

tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa 1560

agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc 1620

tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct 1680

cttcacccca gctgtcctgg aaaagtacaa tgttgttttt aagtacttac tgagtgtgcg 1740

ccgggtgcaa gctgagctgc agcactgctg ggccctacaa atgcagcgca agcacctcaa 1800

gtcgaaccag actgatgcaa tcaagtggcg cctaagaaat cacatggcat ttttggtgga 1860

taatcttcag tactatctcc aggtagatgt gttggagtct cagttctccc agctgcttca 1920

tcagatcaat tctacccgag actttgaaag catccgattg gctcatgacc acttcctgag 1980

caatttgctg gctcaatcct ttatcctatt gaaacctgtg tttcactgcc tgaatgaaat 2040

cctagatctc tgtcacagtt tttgtttgct ggtcagtcag aacctaggcc cactggatga 2100

gcgtggagcc gcccagctga gcattctcgt gaagggcttt agccgccagt cttcactcct 2160

gttcaagatt ctctccagtg ttcggaatca tcagatcaac tcagatttgg ctcaactact 2220

gttacgacta gattataaca aatactatac ccaggctggt ggaactctgg gcagtttcgg 2280

gatgtgaaaa tttctggctc ataaattgaa ataacagcca cgttcccaag gttgtaacag 2340

aagattcaaa acatcccatt ctagccacac acaaataaat atctgcggct tagtgatagg 2400

actctacctt ttctcctaga agcagttact gaacatccag gagtacaact ccttcccatc 2460

attcccatgt ggaagggtct ctcccatcaa ggagaacatg tggcatctct gatcctttac 2520

attgagaaca tttgttggat atgttcattt attcaatagt catttattga gcacctacta 2580

cgtaccttgg tactgttcaa gctgtgggag atacagcggt agacaaacaa tatagagcag 2640

aaagttaaat attttatggt tcatatgtga aaaagtaatt atgtttataa atagactaac 2700

tgctggatgt taccaccaag taagaaagca acaggtaaga taggctttct ctctccctat 2760

accaagtaat ttatacctac acagattggg caattctagc taatgaaaat atacttaaaa 2820
gtatttctta ggccgggcat ggtggctcac acctgtaatc ccagcacttt gggaggccga 2880
ggcgggcgga tcacctgaag tcaggagttt gagaccagcc tgaccaacat gatgaaacct 2940
cgattctact aaaaatacaa aaattagcca ggtgtggtgg catgtgcctg taatcccagc 3000
tactcaggag gctgagacag gagaattgct tgaacctggg aagcagacgc tgcagtgagc 3060
tgagattgtg ccattgcatt ccagcctggg caacaagagc gaaattccgt ctcaaaaaaa 3120
aaaaaaaaaa aaaaagtatt attctccaag aaaaaggtcc ttaagaaaaa attgagatca 3180
agttgttaga tttttaaata ctgaagattg caggcccaat tacccatctt acacaaacca 3240
taggggttga agttatctta atatggccca gccatcactg gtaatcaata ttcatatcag 3300
tgtaagtaaa aagaaatatt cactgaacaa cgccctccaa actgaaaaag aatgcagtgt 3360
tctggcatca ggttatagtc actgcatctg gttttcatca ctacatattc tacacacact 3420
gggaagctct gacaacttat tccctgctat tatcaactaa agatcaccct ttccactgct 3480
gtctctggag caggagctgg caaactatgg cctgctgtct gtttttgtac agttttactg 3540
aaacacagcc gtgcccattt gtttactcat tgtctatggt tgctttcatg ccctcacagc 3600
aaaggcgagt agttgtgatg gatcaaatgg cccacaaagc ctgaaatatt tactctttga 3660
ccctttacag aaaaaaacct tgttgacccc tgctttagag aatgagaagc catgcaggga 3720
tcagtgatgc cagaggaagg gaaggaactg cttccagcta ttgtgacaat aataataata 3780
ataatattgg gtctttgact agaacgtgta acatttccag gtgttctcac ttgtgcttcc 3840
catgtttatc ttacggaagg tcattccatc aagcttatgg tcactgtccc ttcatggcag 3900
ttggtccttt cgttctccct ttagctctaa gagttgggga gtacccacag gtgagctgtg 3960
atctcagctc agagagagag catgaggtct tttttaactg tcaggaaaca gagctgtgcc 4020
caattccact caacttttgg cacaactgtt aatctgggcc ttcacctacc ttaaactgag 4080
tttctgcaag catagcattt tagacaccct ggaataacct tttgggaatg atgccacaga 4140
ataaagttca ctcttaactt ttcaatttcc ttggtgagct gtcctccgta agtgaataag 4200
cctgttgaaa gactcagaga aagtactatg tcttgtcatt tgttctgaga ttaagctcaa 4260
aaaaacagat gaagaaatcc cagttactac aaccaaagag attcaacatt tattttatca 4320
taaaagttca gcaaataaaa ctatatacaa gatccatgca aggaatccag ttacacacaa 4380
gacacattta aaacctggtt aaaacacaat ctccacgata gcagggaata aaaccagtaa 4440
gaccaagtat ctttagtgag aaacataatc gtgtttatat tttggatgct gcttgaatcc 4500
aattctctcc ccaacaatga ggcactggat cacccactct tgtgacacca caggcagctg 4560
caatgcttca gcacacttca gcaccgaggc tgggcatgag ggtccgtca ccaccacatc 4620
aaatacccct aaagcaatat ctgcaaggag caagggaaag tgaagaagga aaggacactc 4680
aacttagccc tccattagaa agagagattt gattctaacc aatacatccc actctgcaca 4740
aaccaaagcc ctattatgtc aaacacactg ctactgatca tgaccaaagg cagagttata 4800
atcactatgt gctgaccttg tagaaatatt taacaaatat acgtccagtg cttcacttat 4860

gttgactcac ctcttgaagg tggtactttt cttctctaag aaacatggat acggtcaacc 4920

tattaggcct gagccttgga ccacaaggcc taacacctac aggtctaagg agatccctgg 4980

aacaaagaca ctacacacac tctttcaggt acctttgtta tgggcacttg aatggtgctg 5040

cttcacagag gctgcaccac cagtcatgag gatctcagac cagagctcca ggaagttctg 5100

ctgttggtct gataccaaga gtaccttcag attctggaaa ggattttcac ggggttgcct 5160

atgaaggaga caggaaagga ccttagcatg acaagtaata tccaacaaac tgcctttctg 5220

caaagggact catgtacatc tgaatgcttt caaaaataaa tgccccatca gacatagtgt 5280

ctcaagcctg taatcccagc actttgggag gctgtcgtgg ttggatctct tgggcctggg 5340

agttcgagac cagcctgggc aatgtggtga gaccccatct ctacaaaaga caacaaaaaa 5400

attagctggg tgtggtggcg agtgcctgta gtcccagcag cttgggaggc tgaggtaggg 5460

ggatcacttc agcctgggag gttgaggctg cagtaagtcg tcactgcgcc actgtactcc 5520

agcctaggtg acagagcaag acttcatctt aaaaaactaa gccctatatt agggtccccc 5580

ttctcttcct tctttctatg aatgatctgt attccttgca ttcctggctt tctaatttcc 5640

atgtttgttc tggggctgag aataatccaa atcatgctcc tgagcctata tatttttaat 5700

gcttgcttaa aacttagttc tctgacttta caggttgaga atattgaacc tatatacaaa 5760

tcttcacaca tttgcaaaag gttcctagcc aatgtaacct agggaaataa actagataaa 5820

ctcctgaagt catttcaaac ccactcaaat ttatcccaca gacattccaa tttctagaaa 5880

gctttactct ctcacctaga ttctcttccc tccaaagctt gctgtcctcc tgcctataca 5940

attctggatg ggcttcaaat acttaccagt ccagaattct ttgctcctca aggctgtacc 6000

cagctggcaa cagataatta cggtagttct ggagctggtt ggcatggcaa ctatcatgga 6060

cccagacatg agacacacaa ggaatcccac tggcaaggca caggaagtac ttccgggttc 6120

gacaatgctg atccgcaatt agaagacact ggtaagctgt gttacactgc aagaaaagaa 6180

gcagagccaa tgggtttggt gacttctgtg gaaagctcct aagcagcagc cataatgagc 6240

catgaagagc agatctgaag actcccaact actacccaaa atgtgattta gtctatcctg 6300

cccaaggcca ctcttctcac tggaaggccc aagtaatttc catagatgtt ctctctgcct 6360

cacctgcagc atactgagga cctaaatcct caacggacaa ccaaaaccta tgaactcagc 6420

ctttcaggct aaaaatcagc aaccctaata ggggtttcta ctactaaaca taaacatcaa 6480

tcttcttttg tcccagcaac agaaccatag ccattaacta acccaaggtc ctaccttctc 6540

ttccctatac acaacaaaaa ttctatttca tgcaaaaaca ttttggcagt ttctcagttc 6600

ctgaaatctc tggctacttt atccaggttc cccaacccct cccaggcctc ttctcaacac 6660

agcaagttgg ctcttatcat tgccactata ttaggttaca caaagaaact cctcacctgg 6720

gcttcattga aatcttcaag gatatagcca gctcctgctc gaagctggga ttctgtatac 6780

tgcttgttga aaggaggaat ttccaaaaat tctatattaa aaaaaaaaac caagata 6837

<210> 17

<211> 733

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 17
cacaatctcc acgatagcag ggaataaaac cagtaagacc aagtatcttt agtgagaaac 60
ataatcgtgt ttatattttg gatgctgctt gaatccaatt ctctccccaa caatgaggca 120
ctggatcacc cactcttgtg acaccacagg cagctgcaat gcttcagcac acttcagcac 180
cgaggctggg catgaggggt ccgtcaccac cacatcaaat acccctaaag caatatctgc 240
aaggagcaag ggaaagtgaa gaaggaaagg acactcaact tagccctcca ttagaaagag 300
agatttgatt ctaaccaata catcccactc tgcacaaacc aaagccctat tatgtcaaac 360
acactgctac tgatcatgac caaaggcaga gttataatca ctatgtgctg accttgtaga 420
aatatttaac aaatatacgt ccagtgcttc acttatgttg actcacctct tgaaggtggt 480
acttttcttc tctaagaaac atggatacgg tcaacctatt aggcctgagc cttggaccac 540
aaggcctaac acctacaggt ctaaggagat ccctggaaca aagacactac acacactctt 600
tcaggtacct ttgttatggg cacttgaatg gtgctgcttc acagaggctg caccaccagt 660
catgaggatc tcagaccaga gctccaggaa gttctgctgt tggtctgata ccaagagtac 720
cttcagattc tgg 733

<210> 18

<211> 734

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 18
gctagaattg cccaatctgt gtaggtataa attacttggt atagggagag agaaagccta 60
tcttacctgt tgctttctta cttggtggta acatccagca gttagtctat ttataaacat 120
aattactttt tcacatatga accataaaat atttaacttt ctgctctata ttgtttgtct 180
accgctgtat ctcccacagc ttgaacagta ccaaggtacg tagtaggtgc tcaataaatg 240
actattgaat aaatgaacat atccaacaaa tgttctcaat gtaaaggatc agagatgcca 300
catgttctcc ttgatgggag agacccttcc acatgggaat gatgggaagg agttgtactc 360
ctggatgttc agtaactgct tctaggagaa aaggtagagt cctatcacta agccgcagat 420
atttatttgt gtgtggctag aatgggatgt tttgaatctt ctgttacaac cttgggaacg 480
tggctgttat ttcaatttat gagccagaaa ttttcacatc ccgaaactgc ccagagttcc 540

accagcctgg gtatagtatt tgttataatc tagtcgtaac agtagttgag ccaaatctga 600

gttgatctga tgattccgaa cactggagag aatcttgaac aggagtgaag actggcggct 660

aaagcccttc acgagaatgc tcagctgggc ggctccacgc tcatccagtg ggcctaggtt 720

ctgactgacc agca 734

<210> 19

<211> 2289

<212> DNA

<213> Homo sapiens

<400> 19

tcgcggccgc gtcracgcgt ggtagggggc ccagagcaag ccgaaggcaa gcacgatggc 60

gctcaccagc cggcccaccc gcgccccgtg ccgcccggag ccccagcggg cgccccgcag 120

ccgtgccagc gtcacgctgt agcagccgag catcagccga aaggaagcac gaaagcggtc 180

agagtctcca ggctcaggtg ggcggcggcg tggaccggcg acgggtggca cagctggcat 240

acgcggtccc tccacaggtg gcggtagacg gcggccggga cggcgagcaa cagggcggcc 300

agccagaccg ccagcagcag gcggcgggcc agggccgggc tgcgcagccg aggcgccagg 360

aaggggcggg tgactgcgag gcagcgctgc aggctgagca ggccggtgag cagcacgctt 420

ggcgtacatg ctgagcgcgc acacgtagta caccgccttg cagcccgcct ggcccagcgg 480

ccaggcctgc cggtcaggaa ggccacaaag agcggcgtga gcagcagcac cgcgccgtcg 540

gccagcgcca ggtgcagcac aagcgtggcc gccagcggtc gcccccgtgc aggctgccag 600

cccgccaagc tccacaccac gaagccgttg ccaggcagcc ccagcagcgc cgccagcagc 660

aggaaggctg tgcctgtggc ccgcgaagtc ttccagctca gcagtgtctc gttccctggg 720

ggacggtagc agaccgacat ccttctgggc ctacaggaca cagaaaaaaa gtggggaagc 780

tgggggaccc tacaaggatc cttggcagga aagcagggat tgtgttcatt ttgagggttt 840

cactgtcagt gagagtctca gcttccatgc aactgtccat cacggctgca actgaaatca 900

gagctgggac acagcgcacc agaagctaaa gtcttgatgc catcaaagga catcccctgc 960

cccattcaca yattcacatc tctgtcacgt ccactaatcg gcaaaaggag aaaagtgaga 1020

gaagatgacc taagtgtgac tgcagcaggc agctctggaa aatgaagcca gagcagtgag 1080

ccagcccctc ctccgaccaa ggaggaagga aagagcagcc ccagcacagg agagaaccac 1140

ccagcccaga agttccaggg aaggaactct ccggtccacc atggagtacc tctcagctct 1200

gaaccccagt gacttactca ggtcagtatc taatataagc tcggagtttg gacggagggt 1260

ctggacctca gctccaccac cccagcgacc tttccgtgtc tgtgatcaca agcggaccat 1320

ccggaaaggc ctgacagctg ccacccgcca ggagctgcta gccaaagcat tggagaccct 1380

actgctgaat ggagtgctaa ccctggtgct agaggaggat ggaactgcag tggacagtga 1440

ggacttcttc cagctgctgg aggatgacac gtgcctgatg gtgttgcagt ctggtcagag 1500

ctggagccct acaaggagtg gagtgctgtc atatgggcct ggacgggaga gccccaagca 1560

cagcaaggac atcggccgat tcacctttga cgtgtacaag caaaaccctc gagacctctt 1620

tggcagcctg aatgtcaaag ccacattcta cgggctctac tctatgagtt gtgactttca 1680

aggacttggc ccaaagaaag tactcaggga gctccttcgt tggacctcca cactgctgca 1740

aggcctgggc catatgttgc tgggaatttc ctccaccctt cgtcatgcag tggagggggc 1800

tgagcagtgg cagcagaagg gccgcctcca ttcctactaa ggggctctga gcttctgccc 1860

ccagaatcat tccaaccgac ccactgcaaa gactatgaca gcatcaaatt tcaggacctg 1920

cagacagtac aggctagata acccacccaa tttccccact gtcctctgat cccctcgtga 1980

cagaaccttt cagcataacg cctcacatcc caagtctata cccttacctg aagaatgctg 2040

ttctttccta gccacctttc tagcctccca cttgccctga aaggccaaga tcaagatgtc 2100

ccccaggcat cttgatccca gcctgactgc tgctacatct aatcccctac caatgcctcc 2160

tgtccctaaa ctccccagca tactgatgac agccctctct gactttacct tgagatctgt 2220

cttcataccc ttcccctcaa actaacaaaa acatttccaa taaaaatatc aaatatttac 2280

cgtcaaccc 2289

<210> 20

<211> 1511

<212> DNA

<213> Homo sapiens

<400> 20

cacatttcat ccttttacat ggttcccatc taccctcaca acacatgtca tcaccaaaga 60

cacacataca agctccaatg gcttttgcca ggcaattctt cctccaggac cccatctggc 120

ccctccctca tccctcccct tggactttgc ccttcttact ggccaggcag gggggccaga 180

gtccaggctt gactcattcc caccttgtcc tgggctgaga tcccaggttt gtaacagaaa 240

acaccactaa agccccagca caggagagaa ccacccagcc cagaagttcc agggaaggaa 300

ctctccggtc caccatggag tacctctcag ctctgaaccc cagtgactta ctcaggtcag 360

tatctaatat aagctcggag tttggacgga gggtctggac ctcagctcca ccaccccagc 420

gacctttccg tgtctgtgat cacaagcgga ccatccggaa aggcctgaca gctgccaccc 480

gccaggagct gctagccaaa gcattggaga ccctactgct gaatggagtg ctaaccctgg 540

tgctagagga ggatggaact gcagtggaca gtgaggactt cttccagctg ctggaggatg 600

acacgtgcct gatggtgttg cagtctggtc agagctggag ccctacaagg agtggagtgc 660

tgtcatatgg cctgggacgg gagaggccca agcacagcaa ggacatcgcc cgattcacct 720

ttgacgtgta caagcaaaac cctcgagacc tctttggcag cctgaatgtc aaagccacat 780

tctacgggct ctactctatg agttgtgact ttcaaggact tggcccaaag aaagtactca 840

gggagctcct tcgttggacc tccacactgc tgcaaggcct gggccatatg ttgctgggaa 900

tttcctccac ccttcgtcat gcagtggagg gggctgagca gtggcagcag aagggccgcc 960

tccattccta ctaaggggct ctgagcttct gcccccagaa tcattccaac cgacccactg 1020

caaagactat gacagcatca aatttcagga cctgcagaca gtacaggcta gataacccac 1080

ccaatttccc cactgtcctc tgatcccctc gtgacagaac ctttcagcat aacgcctcac 1140

atcccaagtc tataccctta cctgaagaat gctgttcttt cctagccacc tttctggcct 1200

cccacttgcc ctgaaaggcc aagatcaaga tgtcccccag gcatcttgat cccagcctga 1260

ctgctgctac atctaatccc ctaccaatgc ctcctgtccc taaactcccc agcatactga 1320

tgacagccct ctctgacttt accttgagat ctgtcttcat acccttcccc tcaaactaac 1380

aaaaacattt ccaataaaaa tatcaaatat ttaccactaa gacttctgac tccaatttaa 1440

accaggaaag ggatggggtg gatacccccat tttgccctcc cccatcaaca cccagtccca 1500

gatccaaagc c 1511

<210> 21

<211> 6530

<212> DNA

<213> Homo sapiens

<400> 21

ttttgttagt ttgaggggaa gggtatgaag acagatctca aggtaaagtc agagagggct 60

gtcatcagta tgctggggag tttagggaca ggaggcattg gtaggggatt agatgtagca 120

gcagtcaggc tgggatcaag atgcctgggg gacatcttga tcttggcctt tcagggcaag 180

tgggaggcca gaaaggtggc taggaaagaa cagcattctt caggtaaggg tatagacttg 240

ggatgtgagg cgttatgctg aaaggttctg tcacgagggg atcagaggac agtggggaaa 300

ttgggtgggt tatctagcct gtactgtctg caggtcctga aatttgatgc tgtcatagtc 360

tttgcagtgg gtcggttgga atgattctgg gggcagaagc tcagagcccc ttagtaggaa 420

tggaggcggc ccttctgctg ccactgctca gcccccctcca ctgcatgacg aagggtggag 480

gaaattccca gcaacatatg gcccaggcct tgcagcagtg tggaggtcca acgaaggagc 540

tccctgaatg gcagagacaa gaggaaatca gatgatttgg aaaacttggg aggaagccat 600

caagctggga gatgaggact ttccacaagc aagagctaac taggggtagg tgggtgcaag 660

aggacgaatt atggggacta tccaactgta ggggatgggg cagtatgaca tgttgatttc 720

tgacctgagt actttctttg ggccaagtcc ttgaaagtca caactcatag agtagagccc 780

gtagaatgtg gctttgacat tcaggctgcc aaagaggtct cgagggtttt gcttgtacac 840

gtcaaaggtg aatcgggcga tgtccttgct gtgcttgggc ctctcccgtc ccaggccata 900

tgacagcact ccactctgta ggacaccctt gtcagtgcag tagatcctca taccagacac 960

ccaccactaa tctccatcag cactgggtca gaccctccct cgcttggact ttctgtccac 1020

tgtgtgacat ccttgacaat tccacaactc ctcctgcacc tggtccccag gatcaggtt 1080

aagctagaga ggaagcccgg gaaagctcta aaggacaggc attggaagca gccccagtat 1140

aggcctctta cccttgtagg gctccagctc tgaccagact gcaacaccat caggcacgtg 1200

tcatcctcca gcagctggaa gaagtcctca ctgtccactg cagttccatc ctcctctagc 1260
accagggtta gcactccatt cagcagtagg gtctccaatg cctgcccaat ggcaagaagc 1320
aagaagggca ggtcttatcc catgcccctt ccctctttag ctgcccaaca tccatcagtt 1380
ggctctagac attggtcgat gtcccacttt gactttccgg cactttgata cctcctaaag 1440
gttgcagctc tccgtgttct tcagtttttg ggggatccta gctagaggct gaccttttc 1500
ctctttgctc ctaccatgtc attggcatct ccccttgctc ccctccaagt cacttctggt 1560
ttggaattgg aaagcaagcc aggttctcac gaagtccacc cttctgtctt atctacaatg 1620
ctgcacctca cttcccacac cctcaagagt tctccagaag tgttttcagt aatagtgttt 1680
aaccttttg agtccttact ctgtgccagg tatgaggact ttacctacat tatcctctta 1740
ctcctttcaa caaccctagg aggtgatgta ttattattgc ctttttatag ttgaagaaac 1800
tgaggttttg gtaggttgaa caacttccca aggtttgaca ggcaggaagt ggcagaatca 1860
gaatttgaac ttgatttgtc acacaaatca cctttccata ctagcttctg aattctgtcc 1920
ctcgaactct ccctatctcc tgctaacccc tgctcccata gaaaagctca ctcggtggaa 1980
aatgaacaaa ttgaccagag ctcattaggc ccactccgct gcttttagcc ctcagaggga 2040
ggggcagctg tgtgacttca gccctctgct ccatcatcac aagttgccac tgttgtggag 2100
ccccttggct acccctgcta taggaaccga ggaacttggc ctacttactt tggctagcag 2160
ctcctggcgg gtggcagctg tcaggccttt ccggatggtc cgcttgtgat cacagacacg 2220
gaaaggtcgc tggggtggtg gagctgaggt ccagaccctc cgtccaaact ccgagcttat 2280
attagatact gacctggtag ttgagaagaa aagtcaagaa ggggcgagga ggggcttggt 2340
gagtgtaaag ggcatgatga gggtagagtg gctagagggc tagggaggga gagatctagg 2400
tttatcgatt agggatgagg gagagaccat ggagtgcagg tgggggcggg tggctcagga 2460
gcttgacaag cccactgtgg agtggggagc aggagaggaa ggggtactgg ttagtctcct 2520
aggggctgag tggagtattg ttgccctgcc tatatcccct aaaggtggag ggtagagcgg 2580
agggttagca gtcacctgag taagtcactg gggttcagag ctgagaggta ctccatggtg 2640
gaccggagag ttccttccct ggaacttctg ggctgggtgg ttctctcctg tgctggggct 2700
ttagtggtgt tttctgttac aaacctggga tctcagccca ggacaaggtg ggaatgagtc 2760
aagcctggac tctggccccc ctgcctggcc agtaagaagg gcaaagtcca aggggaggga 2820
tgagggaggg gccagatggg gtcctggagg aagaattgcc tggcaaaagc cattggagct 2880
tgtatgtgtg tctttggtga tgacatgtgt tgtgagggta gatgggaacc atgtaaaagg 2940
atgaaatgtg acttctggtg ttttttatt tctatggagg gaatttctgg ggacggtttc 3000
tggctctcag gctctgagaa gctgcagttt atgagtggct ctgtgtgtgc tgccacctac 3060
tggagaagcc ataagctgca gctttaggaa aagggaaccc ggggcagagt gtggggaagt 3120
gggatggcag catggcaggg ctttggaaaa tgagaggtga gactgtgtcc aggaagggtg 3180
taaggagagg atggatcctg atacatggat tcaggatcat tagggtcctg tctgggacac 3240
tggccttcct gcttacctgc tctttccttc ctccttggtc ggaggagggg ctggctcact 3300

gctctggctt cattttccag agctgcctgc tgcagtcaca cttaggtcat cttctctcac 3360
ttttctcctt ttgccgatta gtggacgtga cagagatgtg aatggggcag ggatgtcctt 3420
tgatggcatc aagactttag cttctggtgc gctgtgtccc agctctgatt tcagttgcag 3480
ccgtgatgga cagttgcatg gaagctgaga ctctcactga cagtgaaacc ctcaaatgaa 3540
cacaatccct gctttcctgc caaggatcct tgtagggtcc cccagcttcc ccactttttt 3600
tctgtgtcct gacaaagaaa cacagagtaa cttgattgcc ctgtgacctg gccagttgca 3660
tttcccctgc aggcttgagc ccaagccaga gccttgaaaa ggtattcagg ttgttgccca 3720
aaacactgaa aaaaactggc cctggccctg aaccaaatac cttgaaccct cgtaaactcc 3780
ataccctgac cccсttgttt tggatatacc caggtagaac aactctctct cactgtctgt 3840
tgtgaggata cgctgtagcc cactcattaa gtacattctc ctaataaatg ctttggactg 3900
atcaccctgc cagtcttttg tcttgggcaa tctatacttt tctcagaggt tcccaaggcc 3960
tactgaaggg acttaacata ctcttaatgg ctttcctctc tcttgtttta ccttatgccc 4020
tcacttcctg agttaacctc ccaaatacag gatcacctgt acccaagccc ttagctcaag 4080
aatacaggat cacctgtacc caagccctta gctcaagctc tgctttggaa gaacccaaac 4140
taagacagtg ctcctggtgc cctccccaag caacctcaag ttctggctgt tacttgagca 4200
gaggcctttc ttttcccttc ccccagctct atccatctgc caggcccccc tcaaatctct 4260
tcatttccaa gttttgcttg acttttccaa gaggagaggg ctgcttctta gtatgtccct 4320
actcatcctt tcctttcttg tcttgtatcc tggtgcagcc tggtaatggg gcctcttcat 4380
ggttgtgtgt catgactccc taaccattat gcctccatgc atcccctgtt cctcctggaa 4440
cctagcacca tgccttacat ggaaaagctg tcattgacag cccggtgaga gccctgaggg 4500
tggagtgact ggggcagggc ctgaggcaag aggtgggagg aggtaggagg ccaggggctc 4560
agccggacca ggagactgga aacaggcaag gataaggcag gtgggggact gagttgtttg 4620
ggtcacctct gcaggccaga gagaccaggc aacatacaca ctgcagaagg tgggctggga 4680
ggattggggc cagagctggg ggagggatga gaacagaagc aggaccagga ttcagcagag 4740
tcctcctatt tccttccacc accagggaat cttactgccc cacttcagct tgtgctgttt 4800
cctggcaagg caggctctca catgcctgga cgcctgggtg cgttggtgat gggaaggagc 4860
agggtgaggg aggggcccca ggagaggccc aggatgagcc tcatcttgtc cctccccatt 4920
cttgtcttac cctctgcaaa tgtgataggc acaggacagg agtaggcacc tcgcctactg 4980
ctgcttaacc tttcagcttc tccaggcccc caatcctgct tgctcccagc ttggtaagta 5040
gatctgtgca cgtcccttta caccccacca tccagttttg cccagatgtg ctagaatggg 5100
gctggacaaa gaaggagggg ccagactaga ggagtggtgg tagagatagt gacagcctgg 5160
ggtgatgact ttatgcctgt ttaccactga gctctgggaa ggaggccagg agtggggcag 5220
gtcaactgac tgggagcagg ggatctgggt tccaagaagg agttgtgttt gaggtggggt 5280
ctgggtcctc gtggaagtca ggactcccag gcagaaaaga ggcaggctgc agggaagtaa 5340
ggaggaggca tggcaccttc tcatcgggca tcacaggtgg ggttttgccc cacccctgaa 5400
cgccctctgt ggcgccttcc acccacctgt aggcccagaa ggatgtcggt ctgctaccgt 5460

cccccaggga acgagacact gctgagctgg aagacttcgc gggccacagg cacagccttc 5520

ctgctgctgg cggcgctgct ggggctgcct ggcaacggct tcgtggtgtg gagcttggcg 5580

ggctggcggc ctgcacgggg gcgaccgctg gcggccacgc ttgtgctgca cctggcgctg 5640

gccgacggcg cggtgctgct gctcacgccg ctctttgtgg ccttcctgac ccggcaggcc 5700

tggccgctgg gccaggcggg ctgcaaggcg gtgtactacg tgtgcgcgct cagcatgtac 5760

gccagcgtgc tgctcaccgg cctgctcagc ctgcagcgct gcctcgcagt cacccgcccc 5820

ttcctggcgc ctcggctgcg cagcccggcc ctggcccgcc gcctgctgct ggcggtctgg 5880

ctggccgccc tgttgctcgc cgtcccggcc gccgtctacc gccacctgtg gagggaccgc 5940

gtatgccagc tgtgccaccc gtcgccggtc cacgccgccg cccacctgag cctggagact 6000

ctgaccgctt tcgtgcttcc tttcgggctg atgctcggct gctacagcgt gacgctggca 6060

cggctgcggg gcgcccgctg gggctccggg cggcacgggg cgcgggtggg ccggctggtg 6120

agcgccatcg tgcttgcctt cggcttgctc tgggcccccct accacgcagt caaccttctg 6180

caggcggtcg cagcgctggc tccaccggaa ggggccttgg cgaagctggg cggagccggc 6240

caggcggcgc gagcgggaac tacggccttg gccttcttca gttctagcgt caacccggtg 6300

ctctacgtct tcaccgctgg agatctgctg ccccgggcag gtccccgttt cctcacgcgg 6360

ctcttcgaag gctctgggga ggcccgaggg ggcggccgct ctagggaagg gaccatggag 6420

ctccgaacta cccctcagct gaaagtggtg gggcagggcc gcggcaatgg agacccgggg 6480

ggtgggatgg agaaggacgg tccggaatgg gacctttgac agcagaccct 6530

<210> 22

<211> 424

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 22

ggattagatg tagcagcagt caggctggga tcaagatgcc tgggggacat cttgatcttg 60

gcctttcagg gcaagtggga ggctagaaag gtggctagga aagaacagca ttcttcaggt 120

aagggtatag acttgggatg tgaggcgtta tgctgaaagg ttctgtcacg aggggatcag 180

aggacagtgg ggaaattggg tgggttatct agcctgtact gtctgcaggt cctgaaattt 240

gatgctgtca tagtctttgc agtgggtcgg ttggaatgat tctgggggca gaagctcaga 300

gccccttagt aggaatggag gcggcccttc tgctgccact gctcagcccc ctccactgca 360

tgacgaaggg tggaggaaat tcccagcaac atatggccca ggccttgcag cagtgtggag 420

gtcc 424

<210> 23

<211> 424

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 23
ggacctccac actgctgcaa ggcctgggcc atatgttgct gggaatttcc tccacccttc 60

gtcatgcagt ggaggggct gagcagtggc agcagaaggg ccgcctccat tcctactaag 120

gggctctgag cttctgcccc cagaatcatt ccaaccgacc cactgcaaag actatgacag 180

catcaaattt caggacctgc agacagtaca ggctagataa cccacccaat ttccccactg 240

tcctctgatc ccctcgtgac agaacctttc agcataacgc ctcacatccc aagtctatac 300

ccttacctga agaatgctgt tctttcctag ccacctttct agcctcccac ttgccctgaa 360

aggccaagat caagatgtcc cccaggcatc ttgatcccag cctgactgct gctacatcta 420

atcc 424

<210> 24

<211> 7042

<212> DNA

<213> Homo sapiens

<400> 24
aagaagaggt agcgagtgga cgtgactgct ctatcccggg caaaagggat agaaccagag 60

gtggggagtc tgggcagtcg gcgacccgcg aagacttgag gtgccgcagc ggcatccgga 120

gtagcgccgg gctccctccg gggtgcagcc gccgtcgggg gaagggcgcc acaggccggg 180

aagacctcct ccctttgtgt ccagtagtgg ggtccaccgg agggcggccc gtgggccggg 240

cctcaccgcg gcgctccggg actgtggggt caggctgcgt tgggtggacg cccacctcgc 300

caaccttcgg aggtccctgg gggtcttcgt gcgccccggg gctgcagaga tccaggggag 360

gcgcctgtga ggcccggacc tgccccgggg cgaagggtat gtggcgagac agagccctgc 420

acccctaatt cccggtggaa aactcctgtt gccgtttccc tccaccggcc tggagtctcc 480

cagtcttgtc ccggcagtgc cgccctcccc actaagacct aggcgcaaag gcttggctca 540

tggttgacag ctcagagaga gaaagatctg agggaagatg gatgcaaaag ctcgaaattg 600

tttgcttcaa catagagaag ctctggaaaa ggacatcaag acatcctaca tcatggatca 660

catgattagt gatggatttt taacaatatc agaagaggaa aaagtaagaa atgagcccac 720

tcaacagcaa agagcagcta tgctgattaa aatgatactt aaaaaagata atgattccta 780

cgtatcattc tacaatgctc tactacatga aggatataaa gatcttgctg cccttctcca 840

tgatggcatt cctgttgtct cttcttccag tgtaaggaca gtcctgtgtg aaggtggagt 900

accacagagg ccagttgttt ttgtcacaag gaagaagctg gtgaatgcaa ttcagcagaa 960
gctctccaaa ttgaaaggtg aaccaggatg ggtcaccata catggaatgg caggctgtgg 1020
gaagtctgta ttagctgcag aagctgttag agatcattcc cttttagaag gttgtttccc 1080
agggggagtg cattgggttt cagttgggaa acaagacaaa tctgggcttc tgatgaaact 1140
gcagaatctt tgcacacggt tggatcagga tgagagtttt tcccagaggc ttccacttaa 1200
tattgaagag gctaaagacc gtctccgcat tctgatgctt cgcaaacacc caaggtctct 1260
cttgatcttg gatgatgttt gggactcttg ggtgttgaaa gcttttgaca gtcagtgtca 1320
gattcttctt acaaccagag acaagagtgt tacagattca gtaatgggtc ctaaatatgt 1380
agtccctgtg gagagttcct taggaaagga aaaaggactt gaaattttat cccttttgt 1440
taatatgaag aaggcagatt tgccagaaca agctcatagt attataaaag aatgtaaagg 1500
ctctcccctt gtagtatctt taattggtgc acttttacgt gattttccca atcgctggga 1560
gtactacctc aaacagcttc agaataagca gtttaagaga ataaggaaat cttcgtctta 1620
tgattatgag gctctagatg aagccatgtc tataagtgtt gaaatgctca gagaagacat 1680
caaagattat tacacagatc tttccatcct tcagaaggac gttaaggtgc ctacaaaggt 1740
gttatgtatt ctctgggaca tggaaactga agaagttgaa gacatactgc aggagtttgt 1800
aaataagtct cttttattct gtgatcggaa tggaaagtcg tttcgttatt atttacatga 1860
tcttcaagta gattttctta cagagaagaa ttgcagccag cttcaggatc tacataagaa 1920
gataatcact cagtttcaga gatatcacca gccgcatact ctttcaccag atcaggaaga 1980
ctgtatgtat tggtacaact ttctggccta tcacatggcc agtgccaaga tgcacaagga 2040
actttgtgct ttaatgtttt ccctggattg gattaaagca aaaacagaac ttgtaggccc 2100
tgctcatctg attcatgaat ttgtggaata cagacatata ctagatgaaa aggattgtgc 2160
agtcagtgag aattttcagg agtttttatc tttaaatgga caccttcttg gacgacagcc 2220
atttcctaat attgtacaac tgggtctctg tgagccggaa acttcagaag tttatcagca 2280
agctaagctg caggccaagc aggaggtcga taatggaatg ctttacctgg aatggataaa 2340
caaaaaaaac atcacgaatc tttcccgctt agttgtccgc ccccacacag atgctgttta 2400
ccatgcctgc ttttctgagg atggtcagag aatagcttct tgtggagctg ataaaacctt 2460
acaggtgttc aaagctgaaa caggagagaa acttctagaa atcaaggctc atgaggatga 2520
agtgctttgt tgtgcattct ctacagatga cagatttata gcaacctgct cagtggataa 2580
aaaagtgaag atttggaatt ctatgactgg ggaactagta cacacctatg atgagcactc 2640
agagcaagtc aattgctgcc atttcaccaa cagtagtcat catcttctct tagccactgg 2700
gtcaagtgac tgcttcctca aactttggga tttgaatcaa aaagaatgtc gaaataccat 2760
gtttggtcat acaaattcag tcaatcactg cagattttca ccagatgata agcttttggc 2820
tagttgttca gctgatggaa ccttaaagct ttgggatgcg acatcagcaa atgagaggaa 2880
aagcattaat gtgaaacagt tcttcctaaa tttggaggac cctcaagagg atatggaagt 2940
gatagtgaag tgttgttcgt ggtctgctga tggtgcaagg ataatggtgg cagcaaaaaa 3000

taaaatcttt ttgtggaata cagactcacg ttcaaaggtg gctgattgca gaggacattt 3060
aagttgggtt catggtgtga tgttttctcc tgatggatca tcatttttga catcttctga 3120
tgaccagaca atcaggctct gggagacaaa gaaagtatgt aagaactctg ctgtaatgtt 3180
aaagcaagaa gtagatgttg tgtttcaaga aaatgaagtg atggtccttg cagttgacca 3240
tataagacgt ctgcaactca ttaatggaag aacaggtcag attgattatc tgactgaagc 3300
tcaagttagc tgctgttgct taagtccaca tcttcagtac attgcatttg gagatgaaaa 3360
tggagccatt gagattttag aacttgtaaa caatagaatc ttccagtcca ggtttcagca 3420
caagaaaact gtatggcaca tccagttcac agccgatgag aagactctta tttcaagttc 3480
tgatgatgct gaaattcagg tatggaattg gcaattggac aaatgtatct ttctacgagg 3540
ccatcaggaa acagtgaaag actttagact cttgaaaaat tcaagactgc tttcttggtc 3600
atttgatgga acagtgaagg tatggaatat tattactgga aataaagaaa aagactttgt 3660
ctgtcaccag ggtacagtac tttcttgtga catttctcac gatgctacca agttttcatc 3720
tacctctgct gacaagactg caaagatctg gagttttgat ctccttttgc cacttcatga 3780
attgaggggc cacaacggct gtgtgcgctg ctctgccttc tctgtggaca gtaccctgct 3840
ggcaacggga gatgacaatg gagaaatcag gatatggaat gtctcaaacg gtgagcttct 3900
tcatttgtgt gctccgcttt cagaagaagg agctgctacc catggaggct gggtgactga 3960
cctttgcttt tctccagatg gcaaaatgct tatctctgct ggaggatata ttaagtggtg 4020
gaacgttgtc actggggaat cctcacagac cttctacaca aatggaacca atcttaagaa 4080
aatacacgtg tcccctgact tcaaaacata tgtgactgtg gataatcttg gtattttata 4140
tattttacag actttagaat aaaatagtta agcattaatg tagttgaact ttttaaattt 4200
ttgaattgga aaaaaattct aatgaaaccc tgatatcaac tttttataaa gctcttaatt 4260
gttgtgcagt attgcattca ttacaaaagt gtttgtggtt ggatgaataa tattaatgta 4320
gctttttccc aaatgaacat acctttaatc ttgtttttca tgatcatcat taacagtttg 4380
tccttaggat gcaaatgaaa atgtgaatac ataccttgtt gtactgttgg taaaattctg 4440
tcttgatgca ttcaaaatgg ttgacataat taatgagaag aatttggaag aaattggtat 4500
tttaatactg tctgtattta ttactgttat gcaggctgtg cctcagggta gcagtggcct 4560
gctttttgaa ccacacttac cccaaggggg ttttgttctc ctaaatacaa tcttagaggt 4620
tttttgcact ctttaaattt gctttaaaaa tattgtgtct gtgtgcatag tctgcagcat 4680
ttcctttaat tgactcaata agtgagtctt ggatttagca ggcccccccca cctttttttt 4740
ttgtttttgg agacagagtc ttgctttgtt gccaggctgg agtgcagtgg cgcgatctcg 4800
gctcaccaca atcgctgcct cctgggttca agcaattctc ctgcctcagc ctcccgagta 4860
gctgggacta caggtgtgcg cacatgccag gctaattttt gtatttttag tagagacggg 4920
gtttcaccat gttggccggg atggtctcga tctcttgacc tcatgatcta cccgccttgg 4980
cctcccaaag tgctgagatt acaggcgtga gccaccgtgc ctggccaggc cccttctctt 5040
ttaatggaga cagggtcttg cactatcacc caggctggag tgcagtggca taatcatacc 5100
tcattgcagc ctcagactcc tgggttcaag caatcctcct gcctcagcct cccaagtagc 5160

tgagactgca ggcacgagcc accacaccca gctaattttt aagttttctt gtagagacag 5220

ggtctcacta tgttgtctag gctggtcttg aactcttggc ctcaagtaat cctcctgcct 5280

cagcctccca aagtgttggg attgcagata tgagccactg gcctggcctt cagcagttct 5340

ttttgtgaag taaaacttgt atgttggaaa gagtagattt tattggtcta cccttttctc 5400

actgtagctg ctggcagccc tgtgccatat ctggactcta gttgtcagta tctgagttgg 5460

acactattcc tgctccctct tgtttcttac atatcagact tcttacttga atgaaacctg 5520

atctttccta atcctcactt ttttcttttt taaaaagcag tttctccact gctaaatgtt 5580

agtcattgag gtggggccaa ttttaatcat aagccttaat aagattttc taagaaatgt 5640

gaaatagaac aattttcatc taattccatt tacttttaga tgaatggcat tgtgaatgcc 5700

attcttttaa tgaatttcaa gagaattctc tggttttctg tgtaattcca gatgagtcac 5760

tgtaactcta gaagattaac cttccagcca acctattttc ctttcccttg tctctctcat 5820

cctcttttcc ttccttcttt cctttctctt cttttatctc caaggttaat caggaaaaat 5880

agcttttgac aggggaaaaa actcaataac tagctatttt tgacctcctg atcaggaact 5940

ttagttgaag cgtaaatcta aagaaacatt ttctctgaaa tatattatta agggcaatgg 6000

agataaatta atagtagatg tggttcccag aaaatataat caaaattcaa agattttttt 6060

tgtttctgta actggaacta aatcaaatga ttactagtgt taatagtaga taacttgttt 6120

ttattgttgg tgcatattag tataactgtg gggtaggtcg gggagagggt aagggaatag 6180

atcactcaga tgtattttag ataagctatt tagcctttga tggaatcata aatacagtga 6240

atacaatcct ttgcattgtt aaggaggttt tttgttttta aatggtgggt caaggagcta 6300

gtttacaggc ttactgtgat ttaagcaaat gtgaaaagtg aaaccttaat tttatcaaaa 6360

gaaatttctg taaatggtat gtctccttag aatacccaaa tcataatttt atttgtacac 6420

actgttaggg gctcatctca tgtaggcaga gtataaagta ttaccttttg gaattaaaag 6480

ccactgactg ttataaagta taacaacaca catcaggttt taaaaagcct tgaatggccc 6540

ttgtcttaaa aagaaattag gagccaggtg cggtggcacg tgcctgtagt cccagctcct 6600

tgggaggctg agacaggagg attccttgag ccctggagtt tgagtccagc ctgggtgaca 6660

tagcaagacc ctgtcttaaa agaaaaatgg gaagaaagac aaggtaacat gaagaaagaa 6720

gagataccta gtatgatgga gctgcaaatt tcatggcagt tcatgcagtc ggtcaagagg 6780

aggattttgt tttgtagttt gcagatgagc atttctaaag cattttccct tgctgtattt 6840

ttttgtatta taaattacat tggacttcat atatataatt ttttttaca ttatatgtct 6900

cttgtatgtt ttgaaactct tgtatttatg atatagctta tatgatttt ttgccttggt 6960

atacatttta aaatatgaat ttaaaaaatt tttgtaaaaa taaaattcac aaaattgttt 7020

tgaaaaacaa aaaaaaaaaa aa 7042

<210> 25

<211> 3019

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<220>

<221> misc_feature

<222> (2846)..(2846)

<223> Any nucleotide

<400> 25

```
tttttttttt ttttttgaa aaacattttt ggattgtttc attctttgct tgtcatttat      60
ctgttgatta gaccactaaa gtgaaggatt caagctaaat acatcaacct ttctatttag     120
gctttatcag ctatatgtaa attcaattct atcaaaattt tctgagtgcc tcctcagtgt     180
gtctctctga tggttcctgc ccggtatggc tggcatgaag aagatccacg gacttgcgaa     240
tgctaacgcg gggcttgggg atgggtttgg agggtttgtt ttcaaagctt tctggaagtg     300
tggaggagtg tccccctttt cttgcttgta gtgctagctg gtaagcgact tcgaatgcct     360
gtcccagggt taggatgatt tcataggcta aattcacatc aaaggcagta aacacatgac     420
agtagtggtg attagacttc aaatcttttg tgatataggc aaatgttgag aggtcttctg     480
ggtcctgggc agcacaggag atattacgaa tttcatgctc agcaattatg ttcttatttg     540
ttgcatcaat aaatttgact cctttatatg agacagaaag aataatagta gggaccttct     600
tcatttgctc tgtagacttc tgacagttag cccgcatttt tgcacaagca tcttgggttg     660
attctgtccc cctaagctct tttatcagca tagaacctaa ataaaaagct ttgtaatcac     720
acgactggaa gataagcttt tctgggtgat gctgccagta ctgtaccggg gtagaggctg     780
tggcttcatt cggaggtcgc aaggtaatgg aaggttctcc ccagtctcct gtctgagcca     840
tctgcctctc cagttttgat cggggaatat catcaaagta gttttcattt cttctcctcc     900
ttgcatcgcc ctgcatgata atgtgaggaa cgtctaggga gccaccagtg gtgtaagtgc     960
tttggctaag tgatggagac aactgaggag gagtgtgatt accactgggt tccctgaggg    1020
tgatggaccg agggggcttc tgtgggggat cgtcgtgcag cctgtctccc agagatgcca    1080
aaatacgttt cctgtggcca atcaaattga tttttaaaac attaataagt tcaacctccc    1140
agattttttt caacaggtcc atcgaagtgt agccattaat tagaaaggct ttggtgtagt    1200
cgcccagttc aatggaatcc agccactcag ctacagaggt gggatggtag ccatcatgcc    1260
caatgggtct catctttgga aggagctgga ttgcctgtag aattctttgt ctgtgcccag    1320
aattaaggat tccaatttcc aacaaatcct gatcttccat aacattgctt cccataaact    1380
gcacattgtc aaatccatta gccatcaggt ggttctcgta ctgaggtagc ccaatgcttt    1440
ccaaccattg tcccactgtt tggacagggc atctgggtct tgtggtctca ccattcatct    1500
ctttaagttc gttgttgatt ccaacatcta tggaactcat tattttgtca atttcttccc    1560
```

attccgatgt gaaggatggt gttctttcag aattcccttt agaactgtgt tcagcagtgg 1620
aagattcact ccagttaact cttgatgttt tctcattgga aggataggca atgagatcag 1680
aatcagattt agagacactt tttgacaaat gcatgtcgat caaggcttta ggcaatgacc 1740
ttattctccc cagagtacag gctctctcca caaatccccc tgcgttcata acccactggt 1800
ccccattccg agatccactc ctggttgatc ttgtgccaac aatggtatgg ttttcgagtt 1860
ggttgctttt tttgtgaaaa attgttctac tgaccacttt gggtttaatt ttctttacca 1920
aaggttctga gttattttct attggagact gcttgaaagg caaaggactg tttgccaaac 1980
tgacttcatc ttgtcctttt tcacattgtt ctcttttccc atagagatga aatggatttt 2040
caggggactc acaggctgga gaggatccat ggagcaggcc tgcaaattgc ccaggatcat 2100
attcttttgg gggatcattg tcatcctgtc gggagaggtc atctgtatgg ttagttccct 2160
cattcttgac ttctgtggtt cccacagaag aggtgggtgg actagcagga ggactggcag 2220
taaagcttgt gcaccctgta gatgtgttga tttcaaaata ttcttggttg tgattcattc 2280
ggtgaaaatc cagagaagac acaatggatg ttcgctgttt aggctggggt cgaatgactt 2340
ttacaatatt tttgagggca gtatcagggg atggaggtga acaatcaggt gttgggcctg 2400
ttgagctgtt tctatggtta ctagttcctg gagtagtaac tgctacctca gaggcattat 2460
cagttcttgg ggaaggtgcc cttgcaattt ctaaggagca aggtttcttt gtaacagctg 2520
tgtccatgag atcacacaga aagttctcat tttctgaagg aaatgtatcc agagaagcag 2580
atggtacaat ttccatagtg taatttctct tctttggata ggactcctgg gcaagcatgg 2640
ggaagccaag gttcctacat ccattacacg gagttaatgc ttcccaaagt cctgatggcc 2700
cacacgtatt ttcatcatca tcctcttctt ccacttctcc tggtgacaaa ttgattgtag 2760
atgaggttct tacactctgg cttccatttt tcccaagttc ttcctctgaa atcttgctca 2820
aattatctaa gtagtggtct gatatngtgt ggcacaagtc ttcaaacgaa taatcctttt 2880
cttgacagag ttttatttca tccaagagtt ttgataattc tccagtgacg gtttcacttt 2940
tggtcttttg ggaaggagac tcaacaggag atgaaatgtg tgtttcttgt gttgcatctt 3000
cctgtacagg ctcttcgag 3019

<210> 26

<211> 1752

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 26
agaacgcaga ccagcccaag ctgacagctt gagtatgcct tcttctgctg cctggttttg 60
ggggctgtat gacgtactgg tcggtagtaa agattaatat gtaagaaatg tggagctagg 120

atcaagtcat actccacagc ctgcctggca aactatgttt tacttctgac tttgctctct 180
cgctgagaac attaatctgt caagctggcg ggctcctttg atagcaactt tcccaggggc 240
atgatgtggc aatgccacct ctcagcccag gactaccgct attaccccgt ggacggctac 300
tccctgctta aacgcttccc tcttcatcct cttacaggac ccagatgccc tgtccaaaca 360
gtgggacaat ggttggaaag cattgggcta cctcagtacg agaaccacct gatggctaat 420
ggatttgaca atgtgcagtt tatgggaagc aatgttatgg aagatcagga tttgttggaa 480
attggaatcc ttaattctgg gcacagacaa agaattctac aggcaatcca gctccttcca 540
aagatgagac ccattgggca tgatggctac catcccacct ctgtagctga gtggctggat 600
tccattgaac tgggcgacta caccaaagcc tttctaatta atggctacac ttcgatggac 660
ctgttgaaaa aaatctggga ggttgaactt attaatgttt taaaaatcaa tttgattggc 720
cacaggaaac gtattttggc atctctggga gacaggctgc acgacgatcc cccacagaag 780
ccccctcggt ccatcaccct caggacagga gactggggag aaccttccat taccttgcga 840
cctccgaatg aagccacagc ctctaccccg gtacagtact ggcagcatca cccagaaaag 900
cttatcttcc agtcgtgtga ttacaaagct ttttatttag gttctatgct gataaaagag 960
cttaggggga cagaatcaac ccaagatgct tgtgcaaaaa tgcgggctaa ctgtcagaag 1020
tctacagagc aaatgaagaa ggtccctact attattcttt ctgtctcata taaaggagtc 1080
aaatttattg atgcaacaaa taagaacata attgctgagc atgaaattcg taatatctcc 1140
tgtgctgccc aggacccaga agacctctca acatttgcct atatcacaaa agatttgaag 1200
tctaatcacc actactgtca tgtgtttact gcctttgatg tgaatttagc ctatgaaatc 1260
atcctaaccc tgggacaggc attcgaagtc gcttaccagc tagcactaca agcaagaaaa 1320
gggggacact cctccacact tccagaaagc tttgaaaaca aaccctccaa acccatcccc 1380
aagccccgcg ttagcattcg caagtccgtg gatcttcttc atgccagcca taccgggcag 1440
gaaccatcag agagacacac tgaggaggca ctcagaaaat tttgatagaa ttgaatttac 1500
atatagctga taaagcctaa atagaaaggt tgatgtattt agcttgaatc cttcacttta 1560
gtggtctaat caacagataa atgacaagca aagaatgaaa caatccaaaa atgtttttca 1620
aaacaatttt gtgaatttta tttttacaaa aattttttaa attcatattt taaaatgtat 1680
accaaggcaa aaaaatcata taagctatat cataaataca agagtttcaa aacatacaag 1740
agacatataa tg 1752

<210> 27

<211> 367

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 27
ccgcgttagc attcgcaagt ccgtggatct tcttcatgcc agccataccg ggcaggaacc 60

atcagagaga cacactgagg aggcactcag aaaattttga tagaattgaa tttacatata 120

gctgataaag cctaaataga aaggttgatg tatttagctt gaatccttca ctttagtggt 180

ctaatcaaca gataaatgac aagcaaagaa tgaaacaatc caaaaatgtt tttcaaaaca 240

attttgtgaa ttttattttt acaaaaattt tttaaattca tattttaaaa tgtataccaa 300

ggcaaaaaaa tcatataagc tatatcataa atacaagagt ttcaaaacat acaagagaca 360

tataatg 367

<210> 28

<211> 367

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 28
cattatatgt ctcttgtatg ttttgaaact cttgtattta tgatatagct tatatgattt 60

ttttgccttg gtatacattt taaaatatga atttaaaaaa tttttgtaaa aataaaattc 120

acaaaattgt tttgaaaaac atttttggat tgtttcattc tttgcttgtc atttatctgt 180

tgattagacc actaaagtga aggattcaag ctaaatacat caacctttct atttaggctt 240

tatcagctat atgtaaattc aattctatca aaattttctg agtgcctcct cagtgtgtct 300

ctctgatggt tcctgcccgg tatggctggc atgaagaaga tccacggact tgcgaatgct 360

aacgcgg 367

<210> 29

<211> 2457

<212> DNA

<213> Homo sapiens

<400> 29
cacgcagcag gatggcaagg gctccgcttg gggtcctgct cctcttgggg cttctcggca 60

ggggtgtggg gaagaacgag gaactgcgtc tttatcacca tctcttcaac aactatgacc 120

caggaagccg gccagtgcgg gagcctgagg atactgtcac catcagcctc aaggtcaccc 180

tgacgaatct catctcactg aatgaaaaag aggagactct caccactagc gtctggattg 240

gaatcgattg gcaggattac cgactcaact acagcaagga cgactttggg ggtatagaaa 300

ccctgcgagt cccttcagaa ctcgtgtggc tgccagagat tgtgctggaa aacaatattg 360

atggccagtt cggagtggcc tacgacgcca acgtgctcgt ctacgagggc ggctccgtga 420

cgtggctgcc tccggccatc taccgcagcg tctgcgcagt ggaggtcacc tacttcccct 480

tcgattggca gaactgttcg cttattttcc gctctcagac gtacaatgcc gaagaggtgg 540

agttcacttt tgccgtagac aacgacggca agaccatcaa caagatcgac atcgacacag 600

aggcctatac tgagaacggc gagtgggcca tcgacttctg cccgggggtg atccgccgcc 660

accacggtgg cgccaccgac ggcccagggg agactgacgt catctactcg ctcatcatcc 720

gccggaagcc gctcttctac gtcattaaca tcatcgtgcc ctgtgtgctc atctcgggcc 780

tggtgctgct cgcctacttc ctgccggcgc aggccggcgg ccagaaatgc acggtctcca 840

tcaacgtcct gctcgcccag accgtcttct tgttcctcat tgcccagaaa atcccagaga 900

cttctctgag cgtgccgctc ctgggcaggt tccttatttt cgtcatggtg gtcgccacgc 960

tcattgtcat gaattgcgtc atcgtgctca acgtgtccca gcggacgccc accacccacg 1020

ccatgtcccc gcggctgcgc cacgttctcc tggagctgct gccgcgcctc ctgggctccc 1080

cgccgccgcc cgaggccccc cgggccgcct cgcccccaag gcgggcgtcg tcggtgggct 1140

tattgctccg cgcggaggag ctgatactga aaaagccacg gagcgagctc gtgtttgagg 1200

ggcagaggca ccggcagggg acctggacgg ctgccttctg ccagagcctg ggcgccgccg 1260

cccccgaggt ccgctgctgt gtggatgccg tgaacttcgt ggccgagagc acgagagatc 1320

aggaggccac cggcgaggaa gtgtccgact gggtgcgcat ggggaatgcc cttgacaaca 1380

tctgcttctg ggccgctctg gtgctcttca gcgtgggctc cagcctcatc ttcctcgggg 1440

cctacttcaa ccgagtgcct gatctcccct acgcgccgtg tatccagcct tagctcgcac 1500

cgacttcaat ttcccaccca tctccagtag gaaattgatt ttgaaaaagt aggctgccgc 1560

caccacggca ttatgatccc ttccccctgc tgatcaatct gcagtttgtg aacttcacaa 1620

gaatggtgtg tgcccgttcc ctggcgtgtg taggcctggc cgcagtccag ggtcagcag 1680

gaggaaaggg ttcacatagg ctctcaggtg ccagtcttcc agaaagcaag gactgcccctt 1740

cattcagcct tgctgacctc ccagcctttc taaggctcag ccccacggga ctctggtggc 1800

tgccagcttg tgagctatct atctatattc atttcatagc caaacaggag accccctttgc 1860

aggacttgca cacagggagg ctgtagccag gaaaccctct tcttccctgg tctggctctg 1920

ctggagcggg tgggaaccaa acaccttcag tgctggtggc cctcaggccc acaggtttaa 1980

ggctgaggct gccctgaccc ttccacagtc atttcttcta ggttttcttg gcccagcact 2040

gcccatccca ccccatgagg ctcactcatt gcagatccca gcccaccctg cccctttctt 2100

ccccaccctg gaggctctct ctgcctagtc tacagtactg acagaaagca aggacatgcg 2160

gcctgcatgg tgggagctgg ttgaattgtc tttattaaca aacaggatat ccaaggccac 2220

tacattgagg agggggggagg ggggagggag gagaagggtt acttgctgct cacactatat 2280

acagatgcaa gcaaggggcg tggagagtga gggctccctg ctccctccct ccaccgggga 2340

agggcatggg ctagaagagg agaggggggt cgggaatggg gggaatgttt tggctgcggg 2400

gtccccccctc cattccctgg agtttggggg aaggggaatc attaaagtgc tttcaga 2457

<210> 30

<211> 4863

<212> DNA

<213> Homo sapiens

<400> 30

```
ggagatagcg cctgtcagtc ggtgggtcgg tcctcgcgcc ggccctcccc ctccccggtc      60
tccgggggag gcgcggtgga gtccgccccc ggggttctcc gatgggggag aagcggcgac     120
ggcggcagtg gagtaaccga gccggagcgt gagcggcccc ggtgccccgt tccccacgga     180
ggccatgggc gacccagccc ccgccccgcag cctggacgac atcgacctgt ccgccctgcg     240
ggaccctgct gggatctttg agcttgtgga ggtggtcggc aatggaacct acggacaggt     300
gtacaagggt cggcatgtca agacggggca gctggctgcc atcaaggtca tggatgtcac     360
ggaggacgag gaggaagaga tcaaacagga gatcaacatg ctgaaaaagt actctcacca     420
ccgcaacatc gccacctact acggagcctt catcaagaag agcccccccgg gaaacgatga     480
ccagctctgg ctggtgatgg agttctgtgg tgctggttca gtgactgacc tggtaaagaa     540
cacaaaaggc aacgccctga aggaggactg tatcgcctat atctgcaggg agatcctcag     600
gggtctggcc catctccatg cccacaaggt gatccatcga gacatcaagg ggcagaatgt     660
gctgctgaca gagaatgctg aggtcaagct agtggatttt ggggtgagtg ctcagctgga     720
ccgcaccgtg ggcagacgga acactttcat tgggactccc tactggatgg ctccagaggt     780
catcgcctgt gatgagaacc ctgatgccac ctatgattac aggagtgata tttggtctct     840
aggaatcaca gccatcgaga tggcagaggg agcccccccct ctgtgtgaca tgcaccccat     900
gcgagccctc ttcctcattc ctcggaaccc tccgcccagg ctcaagtcca agaagtggtc     960
taagaagttc attgacttca ttgacacatg tctcatcaag acttacctga gccgcccacc    1020
cacggagcag ctactgaagt ttcccttcat ccgggaccag cccacggagc ggcaggtccg    1080
catccagctt aaggaccaca ttgaccgatc ccggaagaag cggggtgaga aagaggagac    1140
agaatatgag tacagcggca gcgaggagga agatgacagc catggagagg aaggagagcc    1200
aagctccatc atgaacgtgc ctggagagtc gactctacgc cgggagtttc tccggctcca    1260
gcaggaaaat aagagcaact cagaggcttt aaaacagcag cagcagctgc agcagcagca    1320
gcagcgagac cccgaggcac acatcaaaca cctgctgcac cagcggcagc ggcgcataga    1380
ggagcagaag gaggagcggc gccgcgtgga ggagcaacag cggcgggagc gggagcagcg    1440
gaagctgcag gagaaggagc agcagcggcg gctggaggac atgcaggctc tgcggcggga    1500
ggaggagcgg cggcaggcgg agcgcgagca ggaatacaag cggaagcagc tggaggagca    1560
gcggcagtca gaacgtctcc agaggcagct gcagcaggag catgcctacc tcaagtccct    1620
gcagcagcag caacagcagc agcagcttca gaaacagcag cagcagcagc tcctgcctgg    1680
ggacaggaag cccctgtacc attatggtcg ggcatgaat cccgctgaca aaccagcctg     1740
ggcccgagag gtagaagaga gaacaaggat gaacaagcag cagaactctc ccttggccaa    1800
gagcaagcca ggcagcacgg ggcctgagcc ccccatcccc caggcctccc cagggccccc    1860
```

aggacccctt tcccagactc ctcctatgca gaggccggtg gagccccagg agggaccgca 1920
caagagcctg gtggcacacc gggtcccact gaagccatat gcagcacctg taccccgatc 1980
ccagtccctg caggaccagc ccacccgaaa cctggctgcc ttcccagcct cccatgaccc 2040
cgaccctgcc atccccgcac ccactgccac gcccagtgcc cgaggagctg tcatccgcca 2100
gaattcagac ccccacctctg aaggacctgg ccccagcccg aatcccccag cctgggtccg 2160
cccagataac gaggccccac ccaaggtgcc tcagaggacc tcatctatcg ccactgccct 2220
taacaccagt ggggccggag ggtcccggcc agcccaggca gtccgtgcca gtaaccccga 2280
cctcaggagg agcgaccctg gctgggaacg ctcggacagc gtccttccag cctctcacgg 2340
gcacctcccc caggctggct cactggagcg gaaccgcgtg ggagtctcct ccaaaccgga 2400
cagctcccct gtgctctccc ctgggaataa agccaagccc gacgaccacc gctcacggcc 2460
aggccggccc gcaagctata agcgagcaat tggtgaggac tttgtgttgc tgaaagagcg 2520
gactctggac gaggcccctc ggcctcccaa gaaggccatg gactactcgt cgtccagcga 2580
ggaggtggaa agcagtgagg acgacgagga ggaaggcgaa ggcgggccag cagaggggag 2640
cagagatacc cctgggggcc gcagcgatgg ggatacagac agcgtcagca ccatggtggt 2700
ccacgacgtc gaggagatca ccgggaccca gcccccatac gggggcggca ccatggtggt 2760
ccagcgcacc cctgaagagg agcggaacct gctgcatgct gacagcaatg ggtacacaaa 2820
cctgcctgac gtggtccagc ccagccactc acccaccgag aacagcaaag gccaaagccc 2880
accctcgaag gatgggagtg gtgactacca gtctcgtggg ctggtaaagg cccctggcaa 2940
gagctcgttc acgatgtttg tggatctagg gatctaccag cctggaggca gtggggacag 3000
catccccatc acagccctag tgggtggaga gggcactcgg ctcgaccagc tgcagtacga 3060
cgtgaggaag ggttctgtgg tcaacgtgaa tcccaccaac acccgggccc acagtgagac 3120
ccctgagatc cggaagtaca agaagcgatt caactccgag atcctctgtg cagccctttg 3180
gggggtcaac ctgctggtgg gcacggagaa cgggctgatg ttgctggacc gaagtgggca 3240
gggcaaggtg tatggactca ttgggcggcg acgcttccag cagatggatg tgctggaggg 3300
gctcaacctg ctcatcacca tctcagggaa aaggaacaaa ctgcgggtgt attacctgtc 3360
ctggctccgg aacaagattc tgcacaatga cccagaagtg gagaagaagc agggctggac 3420
caccgtgggg gacatggagg gctgcgggca ctaccgtgtt gtgaaatacg agcggattaa 3480
gttcctggtc atcgccctca agagctccgt ggaggtgtat gcctgggccc ccaaacccta 3540
ccacaaattc atggccttca agtcctttgc cgacctcccc caccgccctc tgctggtcga 3600
cctgacagta gaggaggggc agcggctcaa ggtcatctat ggctccagtg ctggcttcca 3660
tgctgtggat gtcgactcgg ggaacagcta tgacatctac atccctgtgc acatccagag 3720
ccagatcacg ccccatgcca tcatcttcct ccccaacacc gacggcatgg agatgctgct 3780
gtgctacgag gacgagggtg tctacgtcaa cacgtacggg cgcatcatta aggatgtggt 3840
gctgcagtgg ggggagatgc ctacttctgt ggcctacatc tgctccaacc agataatggg 3900
ctggggtgag aaagccattg agatccgctc tgtggagacg ggccacctcg acggggtctt 3960
catgcacaaa cgagctcaga ggctcaagtt cctgtgtgag cggaatgaca aggtgttttt 4020

tgcctcagtc cgctctgggg gcagcagcca agtttacttc atgactctga accgtaactg 4080

catcatgaac tggtgacggg gccctgggct ggggctgtcc cacactggac ccagctctcc 4140

ccctgcagcc aggcttcccg ggccgcccct ctttcccctc cctgggcttt tgcttttact 4200

ggtttgattt cactggagcc tgctgggaac gtgacctctg accccctgatg ctttcgtgat 4260

cacgtgacca tcctcttccc caacatgtcc tcttcccaaa actgtgcctg tccccagctt 4320

ctggggaggg acacagcttc cccttcccag gaattgagtg ggcctagccc ctcccccctt 4380

ttctccattt gagaggagag tgcttggggc ttgaacccct taccccactg ctgctgactg 4440

ggcagggccc tggacccctt tatttgcacg tcaggggagc cggctccccc cttgaatgta 4500

ccagaccctg ggggggtca ctgggcccta gatttttggg gggtcaccag ccactccagg 4560

ggcagggacc atttcttcat tttctgaaag cactttaatg attccccttc ccccaaactc 4620

cagggaatgg agggggacc ccgccagcca aaacattccc cccattcccg accccctct 4680

cctcttctag cccatgccct tccccggtgg agggagggag cagggagccc tcactctcca 4740

cgcccttgc ttgcatctgt atatagtgtg agcagcaagt aacccttctc cctccccccc 4800

cacccctcct caatgtagtg gccttggata tcctgtttgt taataaagac aattcaacca 4860

gct 4863

<210> 31

<211> 283

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 31

agctggttga attgtcttta ttaacaaaca ggatatccaa ggccactaca ttgaggaggg 60

gggagggggg agggaggaga agggttactt gctgctcaca ctatatacag atgcaagcaa 120

ggggcgtgga gagtgagggc tccctgctcc ctccctccac cggggaaggg catgggctag 180

aagaggagag gggggtcggg aatgggggga atgttttggc tgcggggtcc cccctccatt 240

ccctggagtt tgggggaagg ggaatcatta aagtgctttc aga 283

<210> 32

<211> 283

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 32
tctgaaagca ctttaatgat tccccttccc ccaaactcca gggaatggag gggggacccc 60

gcagccaaaa cattcccccc attcccgacc cccctctcct cttctagccc atgcccttcc 120

ccggtggagg gagggagcag ggagccctca ctctccacgc cccttgcttg catctgtata 180

tagtgtgagc agcaagtaac ccttctcctc cctccccccct cccccctcct caatgtagtg 240

gccttggata tcctgtttgt taataaagac aattcaacca gct 283

<210> 33

<211> 2714

<212> DNA

<213> Homo sapiens

<400> 33
ggcacagggc gaggttttat acacctgaaa gaagagaatg tcaagacgaa gtagccgttt 60

acaagctaag cagcagcccc agcccagcca gacggaatcc ccccaagaag cccagataat 120

ccaggccaag aagaggaaaa ctacccagga tgtcaaaaaa agaagagagg aggtcaccaa 180

gaaacatcag tatgaaatta ggaattgttg gccacctgta ttatctgggg ggatcagtcc 240

ttgcattatc attgaaacac ctcacaaaga aataggaaca agtgatttct ccagatttac 300

aaattacaga tttaaaaatc tttttattaa tccttcacct ttgcctgatt taagctgggg 360

atgttcaaaa gaagtctggc taaacatgtt aaaaaaggag agcagatatg ttcatgacaa 420

acattttgaa gttctgcatt ctgacttgga accacagatg aggtccatac ttctagactg 480

gcttttagag gtatgtgaag tatacacact tcatagggaa acattttatc ttgcacaaga 540

ctttttttgat agatttatgt tgacacaaaa ggatataaat aaaaatatgc ttcaactcat 600

tggaattacc tcattattca ttgcttccaa acttgaggaa atctatgctc ctaaactcca 660

agagtttgct tacgtcactg atggtgcttg cagtgaagag gatatcttaa ggatggaact 720

cattatatta aaggctttaa aatgggaact ttgtcctgta acaatcatct cctggctaaa 780

tctctttctc caagttgatg ctcttaaaga tgctcctaaa gttcttctac ctcagtattc 840

tcaggaaaca ttcattcaaa tagctcagct tttagatctg tgtattctag ccattgattc 900

attagagttc cagtacagaa tactgactgc tgctgccttg tgccatttta cctccattga 960

agtggttaag aaagcctcag gtttggagtg ggacagtatt tcagaatgtg tagattggat 1020

ggtacctttt gtcaatgtag taaaaagtac tagtccagtg aagctgaaga cttttaagaa 1080

gattcctatg gaagacagac ataatatcca gacacataca aactatttgg ctatgctgga 1140

ggaagtaaat tacataaaca ccttcagaaa agggggacag ttgtcaccag tgtgcaatgg 1200

aggcattatg acaccaccga agagcactga aaaaccacca ggaaaacact aaagaagata 1260

actaagcaaa caagttggaa ttcaccaaga ttgggtagaa ctggtatcac tgaactacta 1320

aagttttaca gaaagtagtg ctgtgattga ttgccctagc caattcacaa gttacactgc 1380

cattctgatt ttaaaactta caattggcac taaagaatac atttaattat ttcctatgtt 1440

agctgttaaa gaaacagcag gacttgttta caaagatgtc ttcattccca aggttactgg 1500
atagaagcca accacagtct ataccatagc aatgtttttc ctttaatcca gtgttactgt 1560
gtttatcttg ataaactagg aattttgtca ctggagtttt ggactggata agtgctacct 1620
taaagggtat actaagtgat acagtacttt gaatctagtt gttagattct caaaattcct 1680
acactcttga ctagtgcaat ttggttcttg aaaattaaat ttaaacttgt ttacaaaggt 1740
ttagttttgt aataaggtga ctaatttatc tatagctgct atagcaagct attataaaac 1800
ttgaatttct acaaatggtg aaatttaatg ttttttaaac tagtttattt gccttgccat 1860
aacacatttt ttaactaata aggcttagat gaacatggtg ttcaacctgt gctctaaaca 1920
gtgggagtac caaagaaatt ataaacaaga taaatgctgt ggctccttcc taactggggc 1980
tttcttgaca tgtaggttgc ttggtaataa ccttttttgta tatcacaatt tgggtgaaaa 2040
acttaagtac cctttcaaac tatttatatg aggaagtcac tttactactc taagatatcc 2100
ctaaggaatt ttttttttta atttagtgtg actaaggctt tatttatgtt tgtgaaactg 2160
ttaaggtcct ttctaaattc ctccattgtg agataaggac agtgtcaaag tgataaagct 2220
taacacttga cctaaacttc tattttctta aggaagaaga gtattaaata tatactgact 2280
cctagaaatc tatttattaa aaaaagacat gaaaacttgc tgtacatagg ctagctattt 2340
ctaaatattt taaattagct tttctaaaaa aaaaatccag cctcataaag tagattagaa 2400
aactagattg ctagtttatt ttgttatcag atatgtgaat ctcttctccc tttgaagaaa 2460
ctatacattt attgttacgg tatgaagtct tctgtatagt ttgtttttaa actaatattt 2520
gtttcagtat tttgtctgaa aagaaaacac cactaattgt gtacatatgt attatataaa 2580
cttaaccttt taatactgtt tatttttagc ccattgttta aaaaataaaa gttaaaaaaa 2640
tttaactgct taaaagtaaa gttttgccat tgcttggaga aactttttttt tccttctctg 2700
cgctgccagc tgta 2714

<210> 34
<211> 6773
<212> DNA
<213> Homo sapiens

<400> 34
caagcatgtg atgttcttgt accttcttct gatagtacat ctcaacagtt gactccatat 60
agtcaagtcc atatttgttt gagatctggc aactatcagg aggtaataca gattttcatt 120
gaagacaact taaccttgag tttacctgtc cagttccgac agtcagtcct aagagaactc 180
tttaagaaag ctcaacaggg aaatgaagct ctagatgaaa tctgttttaa agtttgtgcc 240
tgtaatacag tccgtgatat actggaaggc agaacaatta gtgttcaatt taaccagcta 300
tttcttagac caaataaaga gaaaatagac tttcttcttg aggtatgttc aagatcagta 360
aatttagaaa aagcttcaga gtctttgaaa ggaaacatgg ctgcttttct aaagaatgtg 420

tgtctggggt tggaagatct gcagtatgtt ttcatgattt cttcacatga gcttttcatt 480
acattgttga aagatgaaga acgaaagcta cttgttgatc agatgaggaa gagatcccct 540
agagtaaatc tgtgcattaa acctgtaact tcattttatg atatcccagc ttcagcaagt 600
gtcaacattg gtcagttaga gcatcaactt atattgtcag tggatccttg gaggattaga 660
caaattttaa ttgaattaca tggtatgact tcagagcgcc agttctggac agtgtctaat 720
aagtgggaag taccttctgt ctatagtggt gttatcctgg gaattaaaga caatttaaca 780
agagatttgg tttatattct tatggccaaa ggtttgcact gcagtactgt taaggacttt 840
tcccatgcta aacagctctt tgctgcttgt ttggagttgg taacagagtt ctcaccgaag 900
cttcgtcagg tcatgctgaa tgagatgttg cttttggata ttcatacaca cgaagctggg 960
acagggcagg caggagagag accgccatcc gaccttataa gtagagtacg aggctatctg 1020
gaaatgaggc ttcctgatat tcctcttcgt caagttatag ctgaggaatg tgttgccttt 1080
atgttaaact ggagagaaaa tgaatacctt acactccaag ttcctgcatt tttgcttcag 1140
agtaatccat atgtaaagct tggacagctt ttagcagcta catgcaaaga acttccaggc 1200
cctaaagaaa gcagacggac tgccaaagac ctttgggaag ttgttgttca aatctgtagt 1260
gtgtccagtc agcacaaacg aggaaatgat ggcagagtta gtttaataaa acagagggaa 1320
tctacgttag gtatcatgta tcggagtgaa ctgctttctt ttatcaaaaa attacgagaa 1380
ccactcgttt tgactattat tttatcactc tttgtgaaac ttcacaatgt tcgggaggac 1440
attgtgaatg atattacagc tgaacacatt tctatttggc catcttccat tcccaacctc 1500
cagtctgtgg actttgaagc tgtggcaatc acagtgaaag agctagttcg atatacactc 1560
agtataaatc caaataacca ttcttggtta attatccagg cagatattta ctttgcaacg 1620
aatcagtatt cagcagctct tcactattac ctccaggcag gagctgtgtg ttctgacttc 1680
tttaacaagg ctgtgccccc tgatgtttat acagaccagg taataaaacg aatgataaaa 1740
tgttgttctt tgctgaattg ccacacacag gtggctattt tatgtcagtt cctcagagaa 1800
attgactaca aaacagcgtt taaatctctg caagaacaaa acagtcatga tgctatggac 1860
tcctactacg actacatatg ggatgttacc attttggaat acttgactta tcttcatcat 1920
aaaagaggag aaacagataa aagacaaatt gcaatcaaag ccatcggcca gacagagttg 1980
aatgcaagca atccagaaga agtgttacag ctggcagcgc agagaaggaa aaaaaagttt 2040
ctccaagcaa tggcaaaact ttacttttaa gcagttaaat ttttttaact tttatttttt 2100
aaacaatggg ctaaaaataa acagtattaa aaggttaagt ttatataata catatgtaca 2160
caattagtgg tgttttcttt tcagacaaaa tactgaaaca aatattagtt taaaaacaaa 2220
ctatacagaa gacttcatac cgtaacaata aatgtatagt ttcttcaaag ggagaagaga 2280
ttcacatatc tgataacaaa ataaactagc aatctagttt tctaatctac tttatgaggc 2340
tggatttttt ttttagaaaa gctaatttaa aatatttaga aatagctagc ctatgtacag 2400
caagttttca tgtctttttt taataaatag atttctagga gtcagtatat atttaatact 2460
cttcttcctt aagaaaatag aagtttaggt caagtgttaa gctttatcac tttgacactg 2520
tccttatctc acaatggagg aatttagaaa ggaccttaac agtttcacaa acataaataa 2580

agccttagtc acactaaatt aaaaaaaaaa attccttagg gatatcttag agtagtaaag 2640
tgacttcctc atataaatag tttgaaaggg tacttaagtt tttcacccaa attgtgatat 2700
acaaaaaggt tattaccaag caacctacat gtcaagaaag ccccagttag gaaggagcca 2760
cagcatttat cttgtttata atttctttgg tactcccact gtttagagca caggttgaac 2820
accatgttca tctaagcctt attagttaaa aaatgtgtta tggcaaggca aataaactag 2880
tttaaaaaac attaaatttc accatttgta gaaattcaag ttttataata gcttgctata 2940
gcagctatag ataaattagt caccttatta caaaactaaa cctttgtaaa caagtttaaa 3000
tttaattttc aagaaccaaa ttgcactagt caagagtgta ggaattttga gaatctaaca 3060
actagattca aagtactgta tcacttagta taccctttaa ggtagcactt atccagtcca 3120
aaactccagt gacaaaattc ctagtttatc aagataaaca cagtaacact ggattaaagg 3180
aaaaacattg ctatggtata gactgtggtt ggcttctatc cagtaacctt gggaatgaag 3240
acatctttgt aaacaagtcc tgctgtttct ttaacagcta acataggaaa taattaaatg 3300
tattctttag tgccaattgt aagttttaaa atcagaatgg cagtgtaact tgtgaattgg 3360
ctagggcaat caatcacagc actactttct gtaaaacttt agtagttcag tgataccagt 3420
tctacccaat cttggtgaat tccaacttgt ttgcttagtt atcttcttta gtgttttcct 3480
ggtggttttt cagtgctctt cggtggtgtc ataatgcctc cattgcacac tggtgacaac 3540
tgtccccctt ttctgaaggt gtttatgtaa tttacttcct cctatacatg ggaagaaatc 3600
atgcactgat ttcataaatc aaagtcaaac cagacttctg ggtacttatt tgagattatt 3660
taggcctaat tttaatagtc tttttatgtc tttgcaagtg tgaagggtca tattctgaaa 3720
gtttctgtaa cgttatatat tttttaaact ctttatctag actgttggca ttgatcttga 3780
gacacttcac aaatcttgct ttgatttcaa agtaatttta ttaacttttc tacattttga 3840
aatcagtgtg ccccttagaa ctttctttcc cctgaaactg cctgaaggag tactctattc 3900
ctaccatcag ttttggtgac ttactagatt cagatagcaa agccaaaaaa ctcacaaaaa 3960
aacataccag catagccaaa tagtttgtat gtgtctggat attatgtctg tcttccatag 4020
gaatcttctt aaaagtcttc agcttcactg gactagtact ttttactaca ttgacaaaag 4080
gtaccatcca atctacacat tctgaaatac tgtcccactc caaacctaga tagatagaaa 4140
aaagttagaa aagcatgaag gttgtacatc agaaactatc ttacatatgt ctgatgtact 4200
tgttgctgtt tttgagatat tttaaaagaa accaaatcat aaccaagaag tttagcatgt 4260
caaaacagat tatcactctc aaactattta catgactatg ttgaagggaa aaaggacttc 4320
agaacttctt aaccagtacc ttctacatat gaaattgaaa tggtcaaatc ccaaagaact 4380
cttaaagcag aactataatg ttgattcatt tcaactgtat ttaaattcca tttggtcttt 4440
ttgttgatac acattcagga ttggaaagta cttctaacag aaagataatt actgaacagc 4500
taattttttt tttgccaaag ttttaaatgc atgtttagca gaatgttaaa gttcagagac 4560
tgtagtccca ttagaagttg tgaaaaggta agaagacaac aaatagagag tcttacctga 4620
ggctttctta accacttcaa tggaggtaaa atggcacaag gcagcagcag tcagtattct 4680

gtactggaac tctaatgaat caatggctag aatacacaga tctaaaagct aacaggaaaa 4740
acaaaagtac aagcaattta ggagaaagat gagtactaaa tgtctcttgc taaaacctta 4800
gggatctaga gataaataag ccaccacccg gccaggcgcg gtggctcacg cctgtaatcc 4860
cagcactttg ggaggccgag gcgggtggat cacaaggtca cgagatcgag accatcctgg 4920
ctaacacggt gaaaccccgc ctctactaaa aatacaaaa aattagctgg gcttggtggc 4980
acgcgcctgt agtcccagct actcgggagg ctgaggcagg agaatggcgt gaacccggga 5040
tgcagagctt gcagtgagca gagatcgcgc cactgcactc cagcctgggc gaaagagcga 5100
gactccgtct caaaaaaaaa aaaaacaagc caccaacctg aaggaagtag acaaggaagg 5160
actgttgcaa tacagtgtga catgtactag caggaagggc acctaatcca gattggaaaa 5220
gatagtgatg gcctcaaatt gccataaatg ggtcttaaaa gataagggag ccaggaagag 5280
taggaggcag agaatgttct aggtataggg acattacttg gaactcagtt cacagttcag 5340
aactcctaag gtgaaaaata aataaggagt accttcattt cttatcaaga aagatgaggg 5400
gtggtggcta gaaagaggca tggtctagat tggatcacaa agggtcttta agaagtcaga 5460
attttatagg ctgattcttg aagctactgg aagattttta aatcaaagtt ccattttaag 5520
aaagatacct tagaatgcag tgaagcagac agactagaag aaaacatgtt tattaagcag 5580
tgagattagt taaaaggctg tataatctag gcaataagag ctgaactagt agcagtggaa 5640
tggtatagtg taaaaggggt agatttcaca gatttgagaa gatacttgtg cagtggaatt 5700
aaacttcaat tctctttgtc ctcattggtc cagaaggtag gagaaatggg agaagagctg 5760
ggaattggaa gtgaaatatt actgttatat acctctagaa agtccacatt gtttatcggc 5820
ttatcaaaga tttaccatca ctatcagaag ggtatagctg cctaggacaa tttgggatgc 5880
taggaattct ggatgaaaaa attaagcttt taataaaaag ttttataaaa taaaccaatt 5940
tcagtatact tagtggttat ccaatttgag tattcataat gtgctagatt taagcaccac 6000
tgcccacaaa ttttaaccta ggtgacttaa taattatccc caaatgtctt ccatatgtta 6060
gattttcaca tcccacatag aataagaggg tagattttct tcacttttgt tatatggcag 6120
atacagcagc cttaagatta cttacgagaa gtaagcaaga aagaatggga tctcctcttt 6180
ttttttttt ttaatttttt gagatggagt cttgctctgt tgctcaggct ggagtgcagt 6240
agtgcgatct cggctcactg caacctccac ctcccaggtt ccagcgattc tcctgcctca 6300
gcctcccaag taacatgttg gctaggctgc ctcagccgcc caaactcctg acctcaagtg 6360
atctgcctgc ctgcctcagc cgcccaaagt gctgagatta gagacctgag ccacagtgcc 6420
cggccagatc ctcctcctcc tctacttact tactttgtta aatatgctag cctggaaaag 6480
tttactttga atttatgttc taaaaaattt ttttaacaaa gtaattttaa ttctgatatt 6540
taacttgata ggcactctgt gtatccaaat gtaaagacat catacagaat aattctatgc 6600
cattataaag cttaaacaca actggcgaaa aaaatgcttt tccccatttt atatcaaaaa 6660
gagatacttt agtttggact cctaaagaat gaaagtactc agaaaagtgt aaggactttg 6720
tttttctaga aatattaagc aacataaaca ctggggacag aactttatgc gtc 6773

<210> 35

<211> 1590

<212> DNA

<213> Mus musculus

<400> 35
ctgagaacca gacatcagga tggcaggggc tctgcttggt gccctgcttc tcctgacact 60

ctttggcaga agccagggaa agaatgaaga gcttagcctg tatcaccatc tcttcgacaa 120

ttatgatcca gaatgccggc cagttaggag acctgaggac actgtcacca tcaccctcaa 180

ggtcacccta accaacctca tctcactgaa cgagaaagaa gaaactctga ccaccagtgt 240

ctggattggc attgactggc acgactatcg gctcaactac agcaaggacg attttgcagg 300

tgtaggaatc ctccgggtcc cttcagaaca tgtatggctg ccagagattg ttctagaaaa 360

caatattgat gggcagtttg gagtggccta cgacagcaat gttctagtct atgagggagg 420

ctatgtgagc tggttgcccc cagccatcta ccgcagcacc tgcgcagtgg aggtcaccta 480

tttccccttt gactggcaga actgctctct catttttcgc tcccagacct acaatgctga 540

ggaggtggag ttcatctttg ccgtggatga cgacggcaat accatcaaca agattgacat 600

tgacacggca gcttttaccg agaatggaga atgggccata gactactgcc caggcatgat 660

tcgccgctat gagggaggtt ccacagaagg tcctggagaa actgacgtca tctatacgct 720

catcatccgc cggaagccgc tttttacgt cattaacatc attgtgcctt gcgtgctcat 780

ttctggcttg gtgctgctcg cttacttcct gcctgcgcag gctggtggcc agaaatgcac 840

ggtctctatc aacgtcctgc tagcccagac tgtcttcttg tttctaattg cccagaaaat 900

tccagagact tctctgagcg tgccactgct gggcaggtat cttatattcg tcatggtggt 960

tgccacgctc attgtcatga attgcgtcat cgtgctcaac gtatctttga ggacgccaac 1020

gactcatgct acatcccctc ggctgcgcca gattttatta gagctgctgc cgcgtctcct 1080

gggctcgagc ccacccccag aggatccccg aactgcctca ccagcgaggc gtgcctcgtc 1140

tgtgggcatt ctgctcagag cggaggagct catcttgaaa aagccgcgga gcgaactcgt 1200

gtttgagggt cagaggcatc ggcacggaac ttggaccgca gccctctgcc agaacctggg 1260

tgctgcagcc ccagaaatcc gctgctgtgt ggatgctgtg aactttgtgg ctgagagcac 1320

aagagaccag gaagccactg gagaggaact gtccgactgg gtgcgtatgg ggaaggccct 1380

ggacaatgtc tgtttttggg cagctttggt gctcttcagc gttggttcta ctctcatctt 1440

ccttgggggt tacttcaacc aagttcctga tctcccttac ccaccgtgca tccaaccatg 1500

agcctgcact ggcacccacc tctcccccac cccccaagaa agagattttg aaaacaggcc 1560

gctgacaata aatctggttt gtgaacttgc 1590

<210> 36

<211> 2227

<212> DNA

<213> Mus musculus

<400> 36
tgtgagcagc aagtagccct tctccctcct gtatcctttc tcaatgtagt ggccttggat 60

atatcccctt tgttaataaa gacaattcaa ccagcttcca ccattttgag atcctactat 120

tgttctctct caatcctgga gagatttgag agttgagaat gcagagggta gaggaaaggc 180

attaggctct gtgaagttac tgtgataata gagacgaagt aaggtggatg aataggccag 240

ggatcagtcc tgacacggta ggaccctttg agaatagttt ttaccagccc cagcagggcc 300

aggccagact tctggcttca gtgtttctat atctgggtct tgtaaaaacc tcattggcta 360

tcaactagat aaacattctt taggttagaa ggagccaaga gcaaaattga accaattgcc 420

tccaagtgcc tgaccaaacc acccacccat cttctacttc cctgaggagt tggacccacc 480

cacatgacca cacaacccct cgggcagttc acaaaccaga tttattgtca gcggcctgtt 540

ttcaaaatct ctttcttggg gggtggggga gaggtgggtg ccagtgcagg ctcatggttg 600

gatgcacggt gggtaaggga gatcaggaac ttggttgaag taaccccсaa ggaagatgag 660

agtagaacca acgctgaaga gcaccaaagc tgcccaaaaa cagacattgt ccagggcctt 720

ccccatacgc acccagtcgg acagttcctg tgagagagag cttagcgagg gaggagcctg 780

gagggcgggg catctagcac tgctccgcct caacctccca acccacctct ccagtggctt 840

cctggtctct tgtgctctca gccacaaagt tcacagcatc cacacagcag cggatttctg 900

gggctgcagc acccaggttc tggcagaggg ctgctgctaa ggcaacagca agcgctaggt 960

cattaaaaga gcgtcctaac ggcgagtgta tgcctttgac ccaagagcag tgcttaccgg 1020

tccaagttcc gtgccgatgc ctctgaccct caaacacgag ttcgctccgc ggctttttca 1080

agatgagctc ctccgctctg agcagaatgc ccacagacga ggcacgcctc gctggtgagg 1140

cagttcgggg atcctctggg ggtgggctcg agcccaggag acgcggcagc agctctaata 1200

aaatctgcag ccgggggcaga gagaggttcc aagcccgctt cccacccctg ggcagtactt 1260

tctccaacca gcgcttacct ggcgcagccg aggggatgta gcatgagtcg ttggcgtcct 1320

caaagatacg ttgagcacga tgacgcaatt catgacaatg agcgtggcaa ccaccatgac 1380

gaatataaga tacctgatat acagaagcct gatgtcacag caccccacaa acaaggcact 1440

agctgccctc tacctcacaa ataccacctc gcacagctgg tggcgttact tcttgatcct 1500

cctcaacgat gccagtattg tcctggccct tctgcatata ccatctgttg cggacatgaa 1560

ggggattccc agcaatttgg acaccctgct gtgggtctac cacttccaca gctccaccga 1620

ggtgagggta ttagaatggc agaatctgga gaggtcccca gctcttcctg ctatggccct 1680

ttccatgtga tcattccact cactaccctt gctcctccag gtggccttac agcctccact 1740

tctatcttcc ctggaacttg ctgtggccgc agctcacgaa tatctggtgc aaaggttcag 1800

agagcttaag tcccaggacc ccctggaatc cgacaagtcg cccacccaga aggccaccct 1860

agggctggtg ctaagagaag ctgcagccag catcatgagc tttggagcca ccttgttaga 1920

ggtgctgctc tgggaggctg agggatggga ataaaagggg gagagggcta ggccaacaaa 1980

agcaaggacc tctagcccat atgccccaat gtagatctcg gccctgtggc tgcagcagga 2040
ggtgcagcga ctggacggcg gcaacgactg cccaggccca gccccagaca ctggggatcc 2100
tggtagggcg ctggcccgtg tagccctggc cgcagggcag gggattcggc aagctggaac 2160
ggcagctggc gcaagtgccc ggtacctgat ccagggggcg tggttgtacc tgtgtggacg 2220
aggtttg 2227

<210> 37
<211> 2472
<212> DNA
<213> Homo sapiens

<400> 37
agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg 60
gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgaccccttc ccagcttccc 120
gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag 180
gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg 240
aggagggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct 300
cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca 360
ggtatcgcag gacttccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg 420
ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt 480
gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct 540
gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga 600
tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct 660
agaccagttc cagcttcttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca 720
aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg gggggttgcc 780
tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catggggtca tgtataaaca 840
gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa 900
acaggggcca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg 960
cattgggggа ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga 1020
agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc 1080
ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt 1140
tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac 1200
ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga 1260
acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga 1320
agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg 1380
agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc 1440

agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga 1500

tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa 1560

agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc 1620

tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct 1680

cttcacccca gctgtcctgg aaaaaaatag acaattttaa aaaccaaaca gaatgggact 1740

gtcttctgca agcctaccta caaacaggta caatgttgtt tttaagtact tactgagtgt 1800

gcgccgggtg caagctgagc tgcagcactg ctgggcccta caaatgcagc gcaagcacct 1860

caagtcgaac cagactgatg caatcaagtg gcgcctaaga aatcacatgg catttttggt 1920

ggataatctt cagtactatc tccaggtaga tgtgttggag tctcagttct cccagctgct 1980

tcatcagatc aattctaccc gagactttga aagcatccga ttggctcatg accacttcct 2040

gagcaatttg ctggctcaat cctttatcct attgaaacct gtgtttcact gcctgaatga 2100

aatcctagat ctctgtcaca gtttttgttc gctggtcagt cagaacctag gcccactgga 2160

tgagcgtgga gccgcccagc tgagcattct cgtgaagggc tttagccgcc agtcttcact 2220

cctgttcaag attctctcca gtgttcggaa tcatcagatc aactcagatt tggctcaact 2280

actgttacga ctagattata acaaatacta tacccaggct ggtggaactc tgggcagttt 2340

cgggatgtga aaatttctgg ctcataaatt gaaataacag ccacgttccc aaggttgtaa 2400

cagaagattc aaaacatccc attctagcca cacacaaata aatatctgcg gcttaaaaaa 2460

aaaaaaaaaa aa 2472

<210> 38

<211> 4165

<212> DNA

<213> Homo sapiens

<400> 38

agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg 60

gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgaccccttc ccagcttccc 120

gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag 180

gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg 240

aggagggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct 300

cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca 360

ggtatcgcag gacttccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg 420

ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt 480

gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct 540

gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga 600

tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct 660

agaccagttc cagcttcttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca 720

aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg gggggttgcc 780
tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catggggtca tgtataaaca 840
gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa 900
acaggggcca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg 960
cattggggga ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga 1020
agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc 1080
ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt 1140
tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac 1200
ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga 1260
acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga 1320
agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg 1380
agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc 1440
agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga 1500
tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa 1560
agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc 1620
tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct 1680
cttcaccccа gctgtcctgg aaaagtacaa tgttgttttt aagtacttac tgagtgtgcg 1740
ccgggtgcaa gctgagctgc agcactgctg ggccctacaa atgcagcgca agcacctcaa 1800
gtcgaaccag actgatgcaa tcaagtggcg cctaagaaat cacatggcat ttttggtgga 1860
taatcttcag tactatctcc aggtagatgt gttggagtct cagttctccc agctgcttca 1920
tcagatcaat tctacccgag actttgaaag catccgattg gctcatgacc acttcctgag 1980
caatttgctg gctcaatcct ttatcctatt gaaacctgtg tttcactgcc tgaatgaaat 2040
cctagatctc tgtcacagtt tttgtttgct ggtcagtcag aacctaggcc cactggatga 2100
gcgtggagcc gcccagctga gcattctcgt gaagggcttt agccgccagt cttcactcct 2160
gttcaagatt ctctccagtg ttcggaatca tcagatcaac tcagatttgg ctcaactact 2220
gttacgacta gattataaca aatactatac ccaggctggt ggaactctgg gcagtttcgg 2280
gatgtgaaaa tttctggctc ataaattgaa ataacagcca cgttcccaag gttgtaacag 2340
aagattcaaa acatcccatt ctagccacac acaaataaat atctgcggct tagtgatagg 2400
actctacctt ttctcctaga agcagttact gaacatccag gagtacaact ccttcccatc 2460
attcccatgt ggaagggtct ctcccatcaa ggagaacatg tggcatctct gatcctttac 2520
attgagaaca tttgttggat atgttcattt attcaatagt catttattga gcacctacta 2580
cgtaccttgg tactgttcaa gctgtgggag atacagcggt agacaaacaa tatagagcag 2640
aaagttaaat attttatggt tcatatgtga aaaagtaatt atgtttataa atagactaac 2700
tgctggatgt taccaccaag taagaaagca acaggtaaga taggctttct ctctccctat 2760
accaagtaat ttatacctac acagattggg caattctagc taatgaaaat atacttaaaa 2820

gtatttctta ggccgggcat ggtggctcac acctgtaatc ccagcacttt gggaggccga 2880

ggcgggcgga tcacctgaag tcaggagttt gagaccagcc tgaccaacat gatgaaacct 2940

cgattctact aaaaatacaa aaattagcca ggtgtggtgg catgtgcctg taatcccagc 3000

tactcaggag gctgagacag gagaattgct tgaacctggg aagcagacgc tgcagtgagc 3060

tgagattgtg ccattgcatt ccagcctggg caacaagagc gaaattccgt ctcaaaaaaa 3120

aaaaaaaaaa aaaaagtatt attctccaag aaaaaggtcc ttaagaaaaa attgagatca 3180

agttgttaga tttttaaata ctgaagattg caggcccaat tacccatctt acacaaacca 3240

taggggttga agttatctta atatggccca gccatcactg gtaatcaata ttcatatcag 3300

tgtaagtaaa aagaaatatt cactgaacaa cgccctccaa actgaaaaag aatgcagtgt 3360

tctggcatca ggttatagtc actgcatctg gttttcatca ctacatattc tacacacact 3420

gggaagctct gacaacttat tccctgctat tatcaactaa agatcaccct ttccactgct 3480

gtctctggag caggagctgg caaactatgg cctgctgtct gtttttgtac agttttactg 3540

aaacacagcc gtgcccattt gtttactcat tgtctatggt tgctttcatg ccctcacagc 3600

aaaggcgagt agttgtgatg gatcaaatgg cccacaaagc ctgaaatatt tactctttga 3660

ccctttacag aaaaaaacct tgttgacccc tgctttagag aatgagaagc catgcaggga 3720

tcagtgatgc cagaggaagg gaaggaactg cttccagcta ttgtgacaat aataataata 3780

ataatattgg gtctttgact agaacgtgta acatttccag gtgttctcac ttgtgcttcc 3840

catgtttatc ttacggaagg tcattccatc aagcttatgg tcactgtccc ttcatggcag 3900

ttggtccttt cgttctccct ttagctctaa gagttgggga gtacccacag gtgagctgtg 3960

atctcagctc agagagagag catgaggtct tttttaactg tcaggaaaca gagctgtgcc 4020

caattccact caacttttgg cacaactgtt aatctgggcc ttcacctacc ttaaactgag 4080

tttctgcaag catagcattt tagacaccct ggaataacct tttgggaatg atgccacaga 4140

ataaagttca ctcttaactt ttcaa 4165

<210> 39

<211> 27

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 39
ggagagaacc acccagccca gaagttc 27

<210> 40

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 40
aggaatggag gcggcccttc tgc 23

<210> 41

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 41
cggaggagct catcttgaaa aag 23

<210> 42

<211> 24

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 42
gatcaggaac ttggttgaag taac 24

<210> 43

<211> 25

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 43
tgtgagcagc aagtaaccct tctcc 25

<210> 44

<211> 793

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 44
acagagttga atgcaagcaa tccagaagaa gtgttacagc tggcagcgca gagaaggaaa 60

aaaaagtttc tccaagcaat ggcaaaactt tacttttaag cagttaaatt tttttaactt 120

ttatttttta aacaatgggc taaaaataaa cagtattaaa aggttaagtt tatataatac 180

atatgtacac aattagtggt gttttctttt cagacaaaat actgaaacaa atattagttt 240

aaaaacaaac tatacagaag acttcatacc gtaacaataa atgtatagtt tcttcaaagg 300

gagaagagat tcacatatct gataacaaaa taaactagca atctagtttt ctaatctact 360

ttatgaggct ggattttttt tttagaaaag ctaatttaaa atatttagaa atagctagcc 420

tatgtacagc aagttttcat gtcttttttt aataaataga tttctaggag tcagtatata 480

tttaatactc ttcttcctta agaaaataga agtttaggtc aagtgttaag ctttatcact 540

ttgacactgt ccttatctca caatggagga atttagaaag gaccttaaca gtttcacaaa 600

cataaataaa gccttagtca cactaaatta aaaaaaaaaa ttccttaggg atatcttaga 660

gtagtaaagt gacttcctca tataaatagt ttgaaagggt acttaagttt ttcacccaaa 720

ttgtgatata caaaaaggtt attaccaagc aacctacatg tcaagaaagc cccagttagg 780

aaggagccac agc 793

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL02/00904

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68
US CL : 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST, MEDLINE, BIOTECHNO, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MILNER et al., Selecting Effective Antisense Reagents on Combinatorial Oligonucleotide Arrays. Nature Biotechnology, June 1997, Vol. 15, pages 537-541, see entire document.	1-113
Y	VANHEE-BROSSOLLET et al., Do Natural Antisense Transcripts Make Sense in Eukaryotes? Gene, 1998, Vol. 211, pages 1-9, see entire document.	1-113
Y	KUMAR et al., Antisense RNA: Function and Fate of Duplex RNA in Cells of Higher Eukaryotes. Microbiology and Molecular Biology Reviews, December 1998, Vol. 62, No. 4, pages 1415-1434, see entire document.	1-113

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 March 2003 (27.03.2003)	11 APR 2003

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Felicia D. Roberts for Sean R McGarry
Facsimile No. (703)305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)